

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.**

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1395 ATTCTTTTGAAGAAATT 1414
 |||||
 Db 25015 ATTCTTTTGAAGAAATT 25034

RESULT 45
 AC111865
 LOCUS
 DEFINITION Rattus norvegicus clone CH230-214018, *** SEQUENCING IN PROGRESS
 AC111865 101730 bp DNA linear HTG 13-JUL-2002
 *** 51 unordered pieces.
 AC111865
 VERSION AC111865.2 GI:21736631
 KEYWORDS HTG; HTGS PHASE1.
 SOURCE Norway rat.
 ORGANISM Rattus norvegicus
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.

REFERENCE 1 (bases 1 to 101730)
 AUTHORS Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-Osman, P.R., Allen, C., Alsbrooks, S.L., Amarantunga, H.C., Are, J.R., Ayele, M., Banks, T., Barbatia, J., Benton, J., Bimaga, K., Blankenburg, K., Bonnin, D., Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P., Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Caron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Deigado, O., Denn, A.L., Ding, Y., Dinh, H.H., Delnawate, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hogue, M., Holloway, C., Hollins, B., Honsi, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L., Li, J., Li, Z., Licharge, O., Liu, C., Liu, R., Liu, W., Louissegh, H., Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Maesey, E., Mawhiney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenkwo, S., Ogih, M., Okwuonu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, X., Rives, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shooshitari, N., Sisson, I., Sodergren, E., Sonake, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Tabor, P., Tameria, A., Tameria, K., Tang, H., Tanton, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wang, Q., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D., Weinstein, G., and Gibbs, R.

TITLE Direct Submission
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 101730)
 AUTHORS Worley, K.C.
 TITLE Direct Submission
 JOURNAL Submitted (19-FEB-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
 REFERENCE 3 (bases 1 to 101730)
 AUTHORS Worley, K.C.
 TITLE Direct Submission
 JOURNAL Submitted (13-JUL-2002) Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
 On Jul 12, 2002 this sequence version replaced gi:18701709.
 ----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GOJS
 Center clone name: CH230-214018
 ----- Summary Statistics
 Sequencing vector: Plasmid
 Chemistry: Dye-terminator Big Dye: 100% of reads
 Assembly program: Phrap: version 0.990329
 Consensus quality: 35302 bases at least Q40
 Consensus quality: 38351 bases at least Q30
 Consensus quality: 41540 bases at least Q20

 * NOTE: Estimated insert size may differ from sequence length
 (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 51 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 1006: contig of 1006 bp in length
 1007: gap of unknown length
 1107 2126: contig of 1020 bp in length
 2127 2226: gap of unknown length
 2227 3558: contig of 1332 bp in length
 3559 3658: gap of unknown length
 3659 5301: contig of 1643 bp in length
 5302 5401: gap of unknown length
 5402 6430: contig of 1028 bp in length
 6431 7772: contig of 1243 bp in length
 7773 9466: contig of 1594 bp in length
 9467 9566: gap of unknown length
 9567 10844: contig of 1278 bp in length
 10845 10944: gap of unknown length
 10945 12532: contig of 1588 bp in length
 12533 12632: gap of unknown length
 12633 13916: contig of 1284 bp in length
 13917 14016: gap of unknown length
 14017 15302: contig of 1286 bp in length
 15303 16733: contig of 1331 bp in length
 16734 16833: gap of unknown length
 16834 18532: contig of 1699 bp in length
 18533 18632: gap of unknown length
 18633 20006: contig of 1374 bp in length
 20007 21729: contig of 1623 bp in length
 21730 21829: gap of unknown length
 21830 23555: contig of 1726 bp in length
 23556 23655: gap of unknown length
 23656 24953: contig of 1298 bp in length
 24954 25053: gap of unknown length
 25054 26315: contig of 1262 bp in length
 26316 26415: gap of unknown length
 26416 28020: contig of 1605 bp in length
 28021 28120: gap of unknown length
 28121 29458: contig of 1338 bp in length
 29459 29558: gap of unknown length
 29559 30795: contig of 1236 bp in length
 30796 30894: gap of unknown length
 30895 32472: contig of 1578 bp in length
 32473 32572: gap of unknown length

Search completed: December 17, 2002, 11:28:26
Job time : 8694 secs

* 32573 33967: config of 1395 bp in length
* 33968 34067: gap of unknown length
* 34068 35428: config of 1361 bp in length
* 35429 35528: gap of unknown length
* 35529 36887: config of 1359 bp in length
* 36888 36987: gap of unknown length
* 36988 38655: config of 1878 bp in length
* 38656 38966: gap of unknown length
* 38967 40597: config of 1632 bp in length
* 40598 40697: gap of unknown length
* 40698 42321: config of 1624 bp in length
* 42322 42421: gap of unknown length
* 42422 44358: config of 1947 bp in length
* 44359 44468: gap of unknown length
* 44469 46059: config of 1591 bp in length
* 46060 46159: gap of unknown length
* 46160 47945: config of 1786 bp in length
* 47946 48045: gap of unknown length
* 48046 50298: config of 2253 bp in length
* 50299 50398: gap of unknown length
* 50399 52618: config of 2220 bp in length
* 52719 52718: gap of unknown length
* 52719 54903: config of 2185 bp in length
* 54904 55003: gap of unknown length
* 55004 56112: config of 1109 bp in length
* 56113 56212: gap of unknown length
* 56213 58045: config of 1833 bp in length
* 58046 58145: gap of unknown length
* 58146 59642: config of 1497 bp in length
* 59643 59742: gap of unknown length
* 59743 61011: config of 1269 bp in length
* 61012 61111: gap of unknown length
* 61112 62667: config of 1556 bp in length
* 62668 62767: gap of unknown length
* 62768 64883: config of 2116 bp in length
* 64884 64983: gap of unknown length
* 64984 67363: config of 2380 bp in length
* 67364 67463: gap of unknown length
* 67464 68975: config of 1512 bp in length
* 68976 69075: gap of unknown length
* 69076 71513: config of 2438 bp in length
* 71514 71613: gap of unknown length
* 71614 73900: config of 2287 bp in length
* 73901 74000: gap of unknown length
* 74001 76769: config of 2769 bp in length
* 76770 76869: gap of unknown length
* 76870 79077: config of 2208 bp in length
* 79078 79177: gap of unknown length
* 79178 82287: config of 3110 bp in length
* 82288 82387: gap of unknown length
* 82388 86144: config of 3757 bp in length
* 86145 86244: gap of unknown length
* 86245 91344: config of 5100 bp in length
* 91345 91444: gap of unknown length
* 91445 96049: config of 4605 bp in length
* 96050 96149: gap of unknown length
* 96150 101730: config of 5581 bp in length.

FEATURES
source 1..101730
Location/Qualifiers
1.101730
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/clone="CH230-214018"

BASE COUNT 29228 a 18778 c 18804 g 29744 t 5176 others

Query Match 1.1%: Score 20; DB 2; Length 101730;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1650 ATTTGCAATTATATGACC 1669
|||||
DB 30425 ATTTGCAATTATATGACC 30444

Query Match:	86.78%	Indels:	3
DB:	2	Gaps:	0
US-09-868-987-1 (1-1864) x E72062 (1-1402)			
Qy	2	TGCACTTCGCGCATATTTGTCAGAGGGGATCAGCGCACTGCTAATGGACAATATTTCTGCA	61
Db	817	TrpThrSerAlaIyrCysGlnGluGlyIleSerGlyThrAlaAsnGlyGlnIyrSerAla	836
Qy	62	AACCGTGGATGGCGTATGGCTGTAGTAGATTGACGGTTATATGGTCAGCAGGCCCTATTTTA	121
Db	837	AsnArgGlyTrpArgMetAlaValIleAspGlyTyrMetValSerSerProIleLeu	856
Qy	122	AACGTCCCATTTGAAAATCATGCCAGTGTCTCAGGAAATTTACCCACCGTGAAGTCAGC	181
Db	857	AsnValProLeuIlysAsnHisAlaSerValSerGlyLysPheThrHisArgGluValSer	876
Qy	182	AAACTCGCCTCAGATTTAAAACTCGAGCGATGTCTTTTGTTCGCCAGGTTCTCAGTGAA	241
Db	877	LysLeuAlaSerAspLeuIlysSerGlyAlaMetSerPheValProGluValLeuSerGlu	896
Qy	242	GAGACGATCTCTTCTGATCTTGGGAAAAACAATGTACACAAGGCATTTATCTCAGCATGC	301
Db	897	GluThrIleSerSerAspLeuGlyLysGlnCysThrGlnGlyIleIleSerAlaCys	916
Qy	302	TGTGCGTTCGCAATGCTTATCTTTTGATCAGCGTATATTATAGATTTCGGAGCGTCATC	361
Db	917	CysGlyLeuAlaMetLeuIleValLeuMetSerValTyrArgPheGlyGlyValIle	936
Qy	362	GCTTCGGGAGCTGTCTCTCGAATCTTTTGTCTATCTGGCAGCTCTACAGTATTTCGAT	421
Db	937	AlaSerGlyAlaValLeuLeuAsnLeuLeuLeuIleTrpAlaAlaLeuGlnTyrLeuAsp	956
Qy	422	CGCGCACTCACTTGTTCAGACTCGCTGGGATGTCTTGTCTATGGGATGGCGGTAGAT	481
Db	957	AlaProLeuThrLeuSerGlyLeuAlaGlyIleValLeuAlaMetGlyMetAlaValAsp	976
Qy	482	GCAATGCTCTCTGTTTCGAAAGATCCGAGAGGAATTTTATTGTCCTCAAGCTTAA	541
Db	977	AlaAsnValLeuValPheGluArgIleArgGluGluPheLeuLeuSerGlnSerLeuIys	996
Qy	542	AAATCTGTAGAAAAAGGATATACCAAGGCTTTTGGAGCCATTTTGTATCTTAACTTGACT	601
Db	997	LysSerValGluLysGlyTyrThrLysAlaPheGlyAlaIlePheAspSerAsnLeuThr	1016
Qy	602	ACAGTATTCGCTCAGCACTCTCTTTCTCTAGATACAGGGCCTATTAAAGGTTTGCT	661
Db	1017	ThrValLeuAlaSerAlaLeuLeuPhePheLeuAspThrGlyProIleLysGlyPheAla	1036
Qy	662	TTGACATTTGATTTTAGAATTTTCTCTCAATGTTTACGGCTCTTTTCATCACTAAATTT	721
Db	1037	LeuThrLeuIleLeuGlyIlePheSerSerMetPheThrAlaLeuPheMetThrLysPhe	1056
Qy	722	TTCTTCATGCTGTGGATGAATTAAGACCCCAACATACACAGTTGTCATATGATGAATTC	781
Db	1057	PhePheMetLeuTrpMetAsnLysThrGlnHisThrGlnLeuHisMetMetAsnLysPhe	1076
Qy	782	GTGGGATAAGCATGATNTTCTTGAGAGGATGCCAAAACCTTTGGGCTGTTTCTGGAAGT	841
Db	1077	ValGlyIleLysHisAspPheLeuArgGlyCysLysLysLeuTrpAlaValSerGlySer	1096
Qy	842	GTTTTTCTTTTAGGTTCGCTTCGGGTTTGGAGCCTGGAAATTCGGTTTCGGGAATG	901
Db	1097	ValPheLeuLeuGlyCysValAlaLeuGlyPheGlyAlaTrpAsnSerValLeuGlyMet	1116
Qy	902	GATTTTAAAGAGGGTATGCGTTTACCTTTAATCCAAAAGAGCATGGCATCAGCGATGTT	961
Db	1117	AspPheIysGlyTyrAlaPheThrPheAsnProLysGluHisGlyIleSerAspVal	1136
Qy	962	GCTCAATGCGTCGCAACCTGTGCATAACTACAGGAAGCTGGCTTTCTTCTTAGAGAC	1021
Db	1137	AlaGlnMetArgGlyLysValValHisLysLeuGlnGluAlaGlyLeuSerSerArgAsp	1156

Qy	1022	TTCCGTATTCAACACATTTCGATCTTCAGAAAGATCAAAATCTATTTTAGTGATAAAGCT	1081
Db	1157	PhEargileGlnThrPheGlySerGluLysIleLysIleTyPheSerAspLysAla	1176
Qy	1082	TTAAGCTTATAAGCA-GATACGAGCGCTCTCTCCCTAAAAATTAACGATCATGACGCTGGC	1140
Db	1177	LeuSerTyThrLysAlaAspThrSerLeuSerProLysIleAsnAspHisGluLeuAla	1196
Qy	1141	TTATT-GTGGGATTGTGTGAGAAACAGCGCTTAGATTTCTCTACGGAAACTCTAAACGAA	1199
Db	1197	LeuAlaValGlyLeuLeuSerGluThrGlyLeuAspPheSerThrGluThrLeuAsnGlu	1216
Qy	1200	ACGCAAAA-TTTTGGTCAAGGTAGCAGCAACAACTATCGAAGAAAAATCGGTATCAGCG	1258
Db	1217	ThrGlnAsnPheTrpSerLysValSerSerLysLeuSerLysLysMetArgTy-rGlnAla	1236
Qy	1259	ACCATCGGGCTTTTACGAGCTTTGGCAATCATCTGCTCTATGTGAGTTTCGCGCTTTGAA	1318
Db	1237	ThrIleGlyLeuLeuGlyAlaLeuAlaIleIleLeuLeuTyValSerLeuArgPheGlu	1256
Qy	1319	TGGCAATATGCTTTCAGTCCCGCTATGCGCTTTAAATTCATGACCTTTTGGCTACCTGTGCA	1378
Db	1257	TrpGlnTyAlaPheSerAlaValCysAlaLeuIleHisAspLeuLeuAlaThr-CysAla	1276
Qy	1379	GTCTTGTTTATAGCACATTCTCTTTTGAAGAAAAATTCAAATAGATTTCGAAGCCATTGGT	1438
Db	1277	ValLeuPheIleAlaHisPhePheLeuLysLysIleGlnIleAspLeuGlnAlaIleGly	1296
Qy	1439	GCTTTATGACTCTTATGGGTATTCATTAACATACTTTTGATCATTTTGTATCGTATT	1498
Db	1297	AlaLeuMetThrValLeuGlyTySerLeuAsnAsnThrLeuIleIlePheAspArgIle	1316
Qy	1499	CGTGAAGATCCCAACGCAACCTGTTTACCCCTATATGCATGTTTGTAGTTAAATGATGCCCTT	1558
Db	1317	ArgGluAspArgGlnAlaAsnLeuPheThrProMetHisValLeuValAsnAspAlaLeu	1336
Qy	1559	CAAAAGACGTTACAGCCGCAAGTATGACACAGCTACAACTCTATCAGTTTGTTAATG	1618
Db	1337	GlnLysThrPheSerArgThrValMetThrThrAlaThrThrLeuSerValLeuLeuMet	1356
Qy	1619	CTTTTGTTTATAGCGCGCTCCTCTGCTTTTAAATTTTGCATTTATATGACCATAGGGATT	1678
Db	1357	LeuLeuPheIleGlyGlySerValPheAsnPheAlaPheIleMetThrIleGlyIle	1376
Qy	1679	CTTCTAGAACTTTATCGTCTCTTTATATTATGACCACTCTGTTGTTGTTTATGCTCGGT	1738
Db	1377	LeuLeuGlyThrLeuSerSerLeuTyIleAlaProProLeuLeuLeuPheMetValArg	1396
Qy	1739	AAAGAAAATCGCTCAAAA 1756	
Db	1397	LysGluAsnArgSerLys 1402	
RESULT 3			
A71514			
probable protein export - Chlamydia trachomatis (serotype D, strain UW3/Cx)			
C:Species: Chlamydia trachomatis			
C:Date: 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change 08-Oct-1999			
C:Accession: A71514			
R:Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind,			
Science 282, 754-759, 1998			
A:Title: Genome sequence of an obligate intracellular pathogen of humans: C			
A:Reference number: A71570; MUID:9900809; PMID:9784136			
A:Accession: A71514			
A>Status: preliminary			
A:Molecule type: DNA			
A:Residues: 1-1400 <ARN>			
A:Cross-references: GB:AE001318; GB:AE001273; NID:G3328875; PIDN:AAAC68047.1			
A:Experimental source: serotype D, strain UW-3/Cx			
C:Genetics:			
A:Gene: secD/secF			
Alignment Scores:			
Pred. No.:	1,73e-174	Length:	1400

Alignment Scores: 1.73e-174 Length: 1400
Pred. No.:

Score: 2083.00 Matches: 415
Percent Similarity: 79.28% Conservative: 67
Best Local Similarity: 68.26% Mismatches: 80
Query Match: 62.89% Indels: 46
DB: 2 Gaps: 5

US-09-868-987-1 (1-1864) x A71514 (1-1400)

QY 2 TGGACCTCCGATATTCAGAGGGGATCACGGCCTGCTTAATGACATATTCGCA 61
DB TrpThrSerLysPheCysGlnGluValSerGlyThrLysAsnGlyLeuPheSerGly 831
QY 62 AACGGTATGATGCGCTAGTATGATGATGATGATGATGATGATGATGATGATGAT 121
DB GlyArgGlyTyrPheArgMetAlaValAlaLeuAspGlyTyrValIleSerAspProValLeu 851
QY 122 AACGTCCTTGAATAAATCATGCCAGTCTCAGGGAATTTATCCACCGTGAAGTGAAC 181
DB AsnValProLeuLysAspHisAlaSerValSerGlyAsnPheSerTyrArgGluValHis 871
QY 182 AAATCGCTCAGATTTAAATCTGGAGCATGCTTTTGTCCGAGCTTCTCAGTGA 241
DB ArgLeuAlaAlaAspLeuLysSerGlyAlaMetSerPheIleProGluIleLeuSerGlu 891
QY 242 GAGACGATCTCTTCGATCTGGGAAAAACAATGATACAGAGCATTAATCTCAGCATGC 301
DB GluValValSerProGluLeuGlySerSerGlnArgValGlnGlyLeuSerValVal 911
QY 302 TGTGGCTTGGCAATGCTTATTTGATGAGCGTATATTAATGATTTGAGCGCTCATC 361
DB LeuGlyLeuValValLeuIleValLeuMetSerValTyrTyrArgPheGlyGlyValIle 931
QY 362 GCTTCGGAGCTGTTCTTCTGCAATCTTTGCTTATCTGGGCACTCTCAGATTTGAT 421
DB AlaSerIleAlaValLeuLeuAsnLeuLeuLeuIleTyrPalaSerMetGlnTyrLeuAsp 951
QY 422 GCGCACCTCCTGTCAGGAGCTGCTGGGATGTTCTTGGCATGAGGAGGCTGATAT 481
DB AlaProLeuThrLeuSerGlyLeuAlaGlyIleIleLeuAlaMetGlyMetAlaValAsp 971
QY 482 GCMAATGTTCTTGTATGCAAGAAATCCGAGAGAAATTTTATGTTCTCAAAAGCTTAA 541
DB AlaAsnValLeuValPheGluArgIleArgGluGluTyrLeuLeuThrArgSerLeuSer 991
QY 542 AAATCTGAGAAAAAGATPACCAAGGCTTTGGAGCCATTTTGAATTTCACTTGA 601
DB GluSerValGluAlaGlyTyrLysLysAlaPheSerAlaIlePheAspSerAsnLeuThr 1011
QY 602 ACAGATTTGGCTCGAGCTTCTTCTTCCTAGATACAGGCGCTATTAAGGGTTGCT 661
DB ThrIleLeuAlaSerAlaLeuLeuLeuMetLeuAspThrGlyProIleLysGlyPheAla 1031
QY 662 TTGACATTTGATTTGAGATTTTCTTCAATGTTTACGGCTCTTTGATGACTAAATTT 721
DB LeuThrIleuIleIleGlyIlePheSerSerMetPheThrAlaLeuPheMetThrLysPhe 1051
QY 722 TTTCTCATGCTGTGATGAATPAGACCAACATACAGCTTGATGATGATTAAGTTC 781
DB PhePheValIleTyrValGlnLysThrArgGluThrGlnLeuHisMetMetAsnLysPhe 1071
QY 782 GTGGGATTAAGACATGATTTCTTGTGAGAGATGCAAAAACTTTGGCGTTTCTGAGAGT 841
DB IlegIlyIleLysHisAsnPheLeuLysGluCysLysAspGluTyrValValSerIlyVal 1091
QY 842 GTTTTCTTTTATGTTGGTGTCTCTCGGATTTGAGCGCTGATTCGTTTGGAGATG 901
DB ValLeuValIleGlyCysValGlyLeuGlyPheGlyAlaTyrPheSerValPheIlyMet 1111
QY 902 GATTTTAAAGAGGGTATGCTTTTCACTTTAATCCAAAAAGACATGATGATCGATGTT 961
DB AspPheLysGlyGlyTyrAlaLeuThrLeuAsp-----SerAspVal 1125
QY 962 GCT-----CAATGCGTGGCAAAAGTTTGATTAACATACAGGAAGCTGCT 1006

DB 1126 CysGluTyrAsnProGluGlnMetCysSerValLeuArgLysArgPheGlnGlnIleGly 1145
QY 1007 CTTTCTTCTAGAGACTTCCTGATTTCAAACTTTGGATCTTTCAGAAAAATCAATCTAT 1066
DB LeuSerSerArgAspTyrArgValArgAlaAspSerSerGluLysValLysIleTyr 1165
QY 1067 TTTTATGATTAAGCTTAACTATCTAAGCAGATACGA----- 1105
DB LeuSerGlnAsnAlaLeuAspArgValGluGlnIleGluGlyAlaGlySerGluGlnLys 1185
QY 1106 -----GCTCTCTCTCTTAAATTAAGATCAATGATGATGATGATGATGATGAT 1147
DB GlySerAspTyrHisLeuAlaArgValLeuGlnValLeuSerAspSer----- 1201
QY 1148 GGGATTTGTTGTCAGAAACAGCGCTAGATTTTCTCAGGAAACTCTTAAACGAAACG 1207
DB 1202 -----GlySerSerThrThrSerMetVal 1209
QY 1208 TTT-----TGCTCAAGGTAAGCAAGCAAACTATGCAAGAAATGCGT 1249
DB PheAspAlaSerArgGlySerTyrPheLysValSerGlyLeuSerAsnLysMetArg 1229
QY 1250 TATCAGCGCAGCAATCGGGCTTTTAGAGCTTTGGCATCATCTTGTCTTATGATGATTTG 1309
DB ThrGlnAlaValIleAlaLeuPheGlyAlaLeuGlyIleIleLeuPheTyrValSerLeu 1249
QY 1310 CGCTTGAATGCAATATGCTTTCAGATGCGGATGCGCTTTAATTCATGACCTTTGGCT 1369
DB ArgPheGluTyrPheArgTyrAlaPheSerAlaIleCysAlaLeuMetHisAspLeuVala 1269
QY 1370 ACCGTGACGCTGTTTATAGCACATTTCTTTTGAAGAAATTCAAATAGATTTGCA 1429
DB ThrCysAlaValLeuValAlaLeuHisPhePheLeuGlnArgIleGlnIleAspLeuGln 1289
QY 1430 GCCATTTGCTTATATGACTGATTTTGGGATTTTCAATTAACAATCTTATGATCAATTTT 1489
DB AlaIleGlyAlaLeuMetThrValLeuGlyTyrSerLeuAsnHisThrIleIlePhe 1309
QY 1490 GATGCTATTTGTTGAAGATCGCCAGCAAGCACTGTTTACCCCTATGATGTTTATGTTAAT 1549
DB AspArgIleArgGluAspArgArgGluLysLeuPheThrProMetProIleLeuIleAsn 1329
QY 1550 GATGCCCTTCAAAAAGAGCTTCAGCGCGAGGTATAGCAACAGCTATACCTATGAGTT 1609
DB AspAlaLeuGlnLysThrLeuGlyArgThrValMetThrThrAlaThrThrLeuSerVal 1349
QY 1610 TTGTTAATGCTTTTGTATATAGGCGGCTCCTGCTGCTTATATTTTGCATTTATATGACC 1669
DB LeuValIleLeuLeuPheValGlyGlyGlySerIlePheAsnPheAlaPheIleMetThr 1369
QY 1670 ATAGGAATCTTCTAGGAATTTATGCTGCTTATATTTGACACCACTCTGTTGTTT 1729
DB ValGlyIleLeuLeuGlyThrLeuSerSerLeuTyrIleAlaProLeuLeuPhe 1389
QY 1730 ATGTCCTGTAAGAATTCGCTCA 1753
DB 1390 MetValArgLysGluGluGlnAsn 1397
RESULT 4
A81672
seclP protein, probable TCO733 [imported] - Chlamydia muridarum (strain Nigg)
C/Species: Chlamydia muridarum, Chlamydia trachomatis MoKn
C/Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 11-May-2000
C/Accession: A81672
R/Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey,
C.; Dodson, R.; Gwinn, M.; Nelson, W.; Deboy, R.; Kolonay, J.; McClarty, G.; Salzberg,
Nucleic Acids Res. 28, 1397-1406, 2000
A/Title: Genome sequences of Chlamydia trachomatis MoKn and Chlamydia pneumoniae AR39.
A/Reference number: A81672
A/Accession: A81672
A/Status: preliminary
A/Molecule type: DNA

A;Residues: 1-1400 <TET>

A;Cross-references: GB:AB002341; GB:AB002160; NID:97190754; PIDN:AAF39543.1; PID:9719076

A;Experimental source: strain Nigg (MoPn)

C;Genetics:

A;Gene: TC0733

Alignment Scores:

Pred. No.:	3,58e-173	Length:	1400
Score:	2068.00	Matches:	407
Percent Similarity:	79.61%	Conservative:	77
Best Local Similarity:	66.94%	Mismatches:	78
Query Match:	62.44%	Indels:	46
DB:	2	Gaps:	5

US-09-868-987-1 (1-1864) x AB1672 (1-1400)

```
QY 2 TGGACTTCGCGATATTGTGACGAGGGGATCAGCGCACTGCTAATGGACAATATTCTGCA 61
DB 812 TrpThrSerLysPheCysGlnGluGlyValAsnGlyThrLysAsnSerGlnPheSerGly 831
QY 62 AACCGTGGATCGCGTATGCTGTAGTGATGACGGTATATGGTTCAGCAGCCCTATTTTA 121
DB 832 GlyArgGlyTyrArgMetAlaValValLeuAspGlyTyrValIleSerAspProValLeu 851
QY 122 AACGTCCCAATTGAAAATCATGCCAGTGTCTCAGGGAATTTACCCCGTGAAGTGAGC 181
DB 852 AsnValProLeuLysAspHisAlaSerValSerGlyAsnPheSerTyrArgGluValAsn 871
QY 182 AAACCTCGCTCAGATTTAAATCTGGACGATGCTTTTGTTCGCGAGGTTCTCAGTGAA 241
DB 872 ArgLeuAlaSerAspLeuLysSerGlyAlaMetSerPheIleProGluIleLeuSerGlu 891
QY 242 GAGACGATCTCTCTGATCTTGGGAAAAAACAATGTACACAGGCATTATCTCAGCATGC 301
DB 892 GluValValSerProGluLeuGlyHisSerGlnArgLeuGlnGlyIleLeuSerValPhe 911
QY 302 TGTGGCTGGCAATGCTTATTGTTTGTAGAGCGTATATTATAGATTGGAGCGGTATC 361
DB 912 LeuGlyLeuValValLeuIleAlaLeuMetSerValTyrLysPheGlyGlyValIle 931
QY 362 GCTTCGGGAGCTGTTCTGATCTTCTGATCTTGTCTATCTCGGCGAGCTCTACAGATTG 421
DB 932 AlaSerValAlaValLeuLeuAsnLeuLeuLeuIleThrPheAlaSerMetGlnTyrLeuAsp 951
QY 422 CGCGCACTCACCTGTGTGAGGACTCGCTGGGATGTTCTTGTGTATGGGATGCGCGTAG 481
DB 952 AlaProLeuThrLeuSerGlyLeuAlaGlyIleIleLeuAlaMetGlyMetAlaValAsp 971
QY 482 GCAATGTTCTTGTATTCGAAAGATCCGAGAGGAATTTTATTGTCACAAAGCTTAAA 541
DB 972 AlaAsnValLeuValPheGluArgIleArgGluGluTyrLeuLeuSerArgSerLeuSer 991
QY 542 AAATCTGTAGAAAAGGATATACCAAGGCTTTTGGAGGCATTTTGTATCTTAACCTGACT 601
DB 992 GluSerValGluAlaGlyTyrLysAlaPheSerAlaIlePheAspSerAsnLeuThr 1011
QY 602 ACAGTATTGGCTCAGCACTCTCTTCTCTAGATACAGGCGCTTATAAAGGGTTTCT 661
DB 1012 ThrIleLeuAlaSerAlaLeuLeuLeuMetLeuAspThrGlyProIleLysGlyPheAla 1031
QY 662 TTGACATGATTTTAGGAATTTCTTCTCAATGTTTACGGCTCTTTTCATGACTAAATTT 721
DB 1032 LeuThrLeuIleIleGlyIlePheSerSerMetPheThrSerLeuPheMetThrLysPhe 1051
QY 722 TCTCTCATGCTGTGATGAATAAGCCCAACATACAGATTTGCATATGATGAATGATTC 781
DB 1052 PhePheLeuIleIleThrIleGlnLysThrArgGluThrGlnLeuHisMetMetAsnLysPhe 1071
QY 782 GTGGGGATAAAGCATGATTCTTGTAGGAGTCAAAAAAATCTTGGCGTGTCTTGGAGGT 841
DB 1072 IleGlyValLysHisAsnPheLeuLysGluCysLysArgLeuTrpValValSerGlyAla 1091
QY 842 GTTTTCTTTTAGGTTGGCTGCTCTCGGGTTTGGAGCGCTCGGAATTCGTTTGGGAATG 901
```

RESULT 5

AH2768

protein-export membrane protein SECD [imported] - Agrobacterium tumefaciens (strain C58,
C;Species: Agrobacterium tumefaciens

C;Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 11-Jan-2002

C;Accession: AH2768

```
DB 1092 ValValValLeuGlyCysIleGlyLeuGlyPheGlyAlaTrpAspSerIleLeuGlyMet 1111
QY 902 GATTTTAAAGAGGATGCTGCTTTTACCTTTTAAATCCAAAAGAGCATGCGCATCAGCGATGTT 961
DB 1112 AspPheLysGlyGlyTyrAlaLeuThrLeuAsp-----SerAspVal 1125
QY 962 GCT-----CAAATGCGTGGCAAAAGTTGTGCATAAACTACAGGAAGCTGGT 1006
DB 1126 CysAlaTyrAsnProAspGlnMetCysSerAlaLeuLysLysArgPheArgGluIleGly 1145
QY 1007 CTTTCTTCTAGAGACTTCCGTTATTCAAACTTTCGATCTTCAGAAAAGATCAAAATCTAT 1066
DB 1146 LeuSerSerArgAspTyrArgValArgLysAlaAspSerSerGlyLysValLysIleTyr 1165
QY 1067 TTTAGTCATAAAGCTTTTAAAGCTATATAAGCAGATACGA----- 1105
DB 1166 PheSerGlnAsnAlaLeuAlaArgValGluGlnAlaLysGlyLysGlyTyrSerGluGluAsn 1185
QY 1106 -----GCCTCTCTCTAAATTAACGATCATGAGCTGCGGTTATTGT 1147
DB 1186 GlyAlaAspHisHisLeuAlaGlnValLeuGlnValLeuSerAspSer----- 1201
QY 1148 GGGATTGCTCAGAAAACAGCGCTAGATTCTCTACGAAAACCTCTAAACGAAACGCAAAA 1207
DB 1202 -----SerAspAlaThrSerProAlaAla 1209
QY 1208 TTT-----TGGTCAAAGTAAAGCAGCAAACTATCGAAGAAAATGCGT 1249
DB 1210 PheGluSerSerArgGlySerTrpPheLysValSerGlyGlnLeuSerAsnLysMetArg 1229
QY 1250 TATCAGCGCAGCAATGCTTTTAGGAGCTTTTGGCAATCATCTTGTCTATGTGAGTTG 1309
DB 1230 ThrGlnAlaIleMetAlaLeuPheGlyAlaLeuGlyIleIleLeuLeuTyrValSerLeu 1249
QY 1310 CGCTTGAATGGCAATATGCTTTTTCAGTCCGCTATGCGCTTTAATTCATGACCTTTTGCT 1369
DB 1250 ArgPheGluTrpArgTyrAlaPheSerAlaIleCysAlaLeuMetHisAspLeuLeuAla 1269
QY 1370 ACCTGTGCGAGCTTGTATTATAGCACATCTTTTGAAGAAAATTCAAATAGATTGCAA 1429
DB 1270 ThrCysSerValLeuValAlaLeuHisPhePheLeuHisArgIleGlnIleAspLeuGln 1289
QY 1430 GCCATTGCTGTTTAAATGACTGATTTGCGGTATTTCATTAACAATACCTTTGATCATTTT 1489
DB 1290 AlaValGlyAlaLeuMetThrValLeuGlyTyrSerLeuAsnAsnThrLeuIleIlePhe 1309
QY 1490 GATCGTATTCTGTGAAGATCGCCAAAGCGAACCTGTTTACCCTATGCATGTTTATAGTTAAT 1549
DB 1310 AspArgIleArgGluAspArgArgGluLysLeuPheThrProMetProIleLeuIleAsn 1329
QY 1550 GATGCCCTTCAAAGACGTTTCAGCCGCGATATGACACAGCTACACACTCTATCAGTT 1609
DB 1330 AspAlaLeuGlnLysThrLeuGlyArgThrValMetThrThrAlaThrThrLeuSerVal 1349
QY 1610 TTGTTAATGCTTTTGTATTATAGCGGCTCTCTCTCTTTAATTTTGTATTTATATGACC 1669
DB 1350 LeuValIleLeuLeuPheValGlyGlySerIlePheAsnPheAlaPheIleMetThr 1369
QY 1670 ATAGGGATTTCTTAGGAATTTTATCGTCTCTTTTATATTCACCACTCTCTGTTGTTT 1729
DB 1370 ValGlyIleLeuLeuGlyThrLeuSerSerLeuTyrIleAlaProProLeuLeuLeuPhe 1389
QY 1730 ATGCTCCGTAAGAAAATCGCTCA 1753
DB 1390 MetValArgLysGluGluGlnAsn 1397
```

R:Wood, D.W.; Serubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L.
 erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McCell
 ; Karp, P.; Romero, P.; Zhang, S.
 Science 294, 2317-2323, 2001

A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
 ster, E.W.

A>Title: The Genome of the Natural Genetic Engineer *Agrobacterium tumefaciens* C58.

A/Reference number: AB2577; PMID:11743193

A/Accession: AH2768

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-857 <KUR>

A/Cross-references: GB:AB008688; PIDN:AL42566.1; PID:G17739990; GSPDB:GN00186

A/Experimental source: strain C58 (Dupont)

C/Genetics:

A/Map position: circular chromosome

Alignment Scores:

Score:	2.21e-65	Length:	857
Best Local Similarity:	57.75%	Matches:	198
Query Match:	25.36%	Mismatches:	209
		Indels:	39
		Gaps:	13

US-09-868-987-1 (1-1864) x AH2768 (1-857)

```

QY 7 TTCCGCATATTGTCAGAGGAGATCGCGCATGCTGTAATGACAAATTTCTGCAACCG 66
Db 282 PheargPheaspSerArgGlyAlaGlnArgHe-AlaGlnAlaThr--GlnGlnAsnVa 300
QY 67 TGGATGGCGTATGCTAGTATGATGACGTTATATGTCAGACGCTTATTTAAACGT 126
Db 300 IGLyIleProPheAlaIleValIleuAspAsnGlnValIleSerAlaProValIleArgI 320
QY 127 CCCA---TTGAAAATCATGCCAGTCTCAGGGAAATTTCCACCGTGAAGAGCAA 183
Db 320 uPrlIleIleGlySerGlyGlnIleSerGlySerPheSerValGlnGlyAlaAsnAs 340
QY 184 ACTGCCTCCAGATTTAAATCTGAGAGCATGCTTTTGTCCCGAGCTTCAGTGAAGA 243
Db 340 pLeuAlaValIleuAsnAlaGlyAlaIleuProAlaThrIleuThrValIaGlnGluAr 360
QY 244 GACGATCTCTTCTGATCTTGGGAAAAACAATGACACAGGCAATTCATCAGCATGCTG 303
Db 360 gThrValGlyProSerIleuGlyAsnAspSerIleThrAlaGlyLeuThrAlaSerAlaI 380
QY 304 TGGCTGGCAATGCTTATTTGTTGATGAGCGTATATATGATGATTTGGAGCGTCATCG 363
Db 380 eGlyAlaValaGlyValIleuIlePheMetPheValPheIleGlyPhePheGlyLeuLeuAl 400
QY 364 TTCCGAGCATCTTCTTGAATCTTGTATCTGAGCAGCTCTACAGTATTTGGATGC 423
Db 400 aAsnIleAlaIleuIleValaIleuValaIleuIleuIleAlaIleuSerValIleGlySe 420
QY 424 GCCACTCACCTTGCAGAGCTCGCTGGAGATTGTTCTGTATGGGAGTGGCCGTAGATGC 483
Db 420 rThleuThrIleuProGlyIleAlaGlyIleValIleuThrIleGlyMetAlaValaSpse 440
QY 484 AAAATGTTTGTATTCGAAAGATTCGAGAGGAATTTTATTTGTCGAAAGCTTTAAAAA 543
Db 440 rAsnValaIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 560
QY 544 ATCTGTGAAAAAGATATACAGGCTTTTGGAGCACTTTTGTATTTGATTTGATGAC 603
Db 460 nAlaIleuAspAsnGlyPheThrArgAlaPheAlaThrIleIleAlaIleuIleuIleuTh 480
QY 604 AGATTTGGCTCAGACATCTTTCTTCTCTGATGATACAGGCTCTATTAAGGGTTGCTTT 663
Db 480 rLeuIleValaIleuSerValaIleuPheIleuIleuIleuIleuIleuIleuIleuIleu 500
QY 664 GACATGATTTTAAAGATTTTCTCTCAATGTTTACGGCTCTTTCAATGATTAATTTT 723

```

```

Db 500 lThrIleAlaValaGlyIleIleIleThrIleThrValPheThrAlaIleThleuThrAlaIlePme 520
QY 724 CTTTCAGCTGTGATGATTAAGACCAACATACAGTTGCAT-----ATGATGAA 774
Db 520 tPheGlyIleThrIleValaIleuArgSerAlaProGlySH:IleuProIleGlyValaIleThrAl 540
QY 775 TAAGTTCGTGGGATTAAGATGATTTCTTGAGAGATGCAAAAACTTTGGGCTGTTCC 834
Db 540 aMetPheAspGlyArgAspIleProPheMetAlaGlyArgValaValaPheMetIleTh 560
QY 835 TGGAGTGTCTTTCTTTAGAGTTCGCTTCGCTGCTGGGTTGGAGCCCTGGAATCCGTTT 894
Db 560 rGly---ValIleMetIleuAlaCysValaGly---GlyPheValaIleAlaIleuIleuIle 578
QY 895 GGGATGATTTTAAAGAGGATATGCTTACCTTTAATCAAAAGACATGGCATCAG 954
Db 578 uGlyIleAspPheGlnGlyIleSerValIleGluValaIleGln---GlyGluAl 597
QY 955 CGATGTTGCTCAATAGCTGTGCGCAAGTTGTCATTAACATACAGGAGCTGCTTCTTC 1014
Db 597 aAspIleuAlaAspIleArg-----GluArgIleuAsnGlnIleuAsnIleuGly-- 612
QY 1015 TAGAGACTTCGCTATTCACATTTTGATCTTCAGAAAAGATCAAAATCTATTTAGTGA 1074
Db 613 ---GluIleGlnAlaGlnAsnIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 629
QY 1075 TAAAGCTTTAAGCTATATACAGATACAGAGCTCTCTCTTAATTAACATGATCAG 1134
Db 630 -GlnAlaGlnAspGlyGlyIleuAsnAlaGlnIleuIleuIleuIleuIleuIleuIleuIleu 645
QY 1135 CTGGCGTATTTGGGATGTTGTCAGAAACAGGCTCAGATTCTCTCAGGAAACCTTAA 1194
Db 646 -----ValArgIleGlyIleu-----GluAspIle 653
QY 1195 ACGAAGCAAAATTTTGTCAAGGTACAGCAAACTATCAAGAAATGCGTATCA 1254
Db 653 sTyraAspPheArgArgValaIleuValaIleuIleuIleuIleuIleuIleuIleuIleuIleu 673
QY 1255 GCGCAGCATCGGCTTTAGAGCTTTGGCAATCATCTGCTATGATGATTTGGCGCTT 1314
Db 673 rSerThrIleGlyValIleuIleuAlaIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 693
QY 1315 TGAATGCAATATGCTTTCAGTGCCTGATGCGCTTATATCAIAGCTTTGGCTACTG 1374
Db 693 eGluTrpGlnPheAlaIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 712
QY 1375 TGCAGTCTGTTTATACACATTTCTTTTGAAGAAATTCAAATGATTTGCAAGCAT 1434
Db 713 -----IleGlyIleuPheValaIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 728
QY 1435 TGGCTTTAATGATCTGATTTGGGATTTGATTTAAACATATCTTGATCATTTTGTATG 1494
Db 728 eAlaIleAlaIleuThrIleIleGlyIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 748
QY 1495 TATTCGTAAGATGCGCAAGCAACCTGTTACCCCTATGCAAGCTTTTACTATATATGC 1554
Db 748 gValArgGlnAsnIleuArgArgIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 768
QY 1555 CTTCAAAAGAGCTTCAGCGCAGGATATGATGACAGCAAGCTATACAGTTTGTGT 1614
Db 768 rLeuAsnGlnThrIleuSerArgIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 788
QY 1615 AATGCTTTTGTATTAAGCGCGCTCTCTGCTTATTTTGAATTTGATTTATGACCAATAG 1674
Db 788 uAlaIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 808
QY 1675 GATCTCTTGAAGACTTATGCTCTCTTATATTTGACCAACCTCTGTGTTG---TTAT 1731
Db 808 yValaGlyIleGlyValaPheSerSerValIleIleAlaIleuIleuIleuIleuIleuIleuIleu 828
QY 1732 GGTCCGTAAGAAATGCGC 1750

```

Db 828 sLeuArgProAspSerLys 834
RESULT 6
B97549
probable protein-export membrane protein [imported] - Agrobacterium tumefaciens (strain
C:Species: Agrobacterium tumefaciens
C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 11-Jan-2002
C:Accession: B97549
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A:Reference number: A97359; PMID:11743194
A:Accession: B97549
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-876 <KUR>
A:Cross-references: GB:AE007869; PIDN:AAK87347.1; PID:g15156649; GSPDB:GN00169
C:Genetics:
A:Gene: AGR_C_2877
A:Map position: circular chromosome
Alignment Scores:
Pred. No.: 2, 22e-65 Length: 876
Score: 840.00 Matches: 198
Percent Similarity: 57.75% Conservative: 141
Best Local Similarity: 33.73% Mismatches: 209
Query Match: 25.36% Indels: 39
DB: 2 Gaps: 13
US-09-868-987-1 (1-1864) x B97549 (1-876)
Qy 7 TTCGGCATATTGTCAGGAGGGGATCAGCGGCATCTAATGCAATATTCTGCAAAACCG 66
Db 301 PheArgPheAspSerArgGlyAlaGlnArgPhe-AlaGlnAlaThr---GlnGlnAsnVa 319
Qy 67 TGGATGGCGTATGCTGTAGTGTATGAGCGTTATATGTCAGAGCCCTATTAAACGT 126
Db 319 lGlyLysProPheAlaIleValLeuAsnGlnValIleSerAlaProValIleArgGl 339
Qy 127 CCCA---TTGAAAATCATCGCAGTGTCTCAGGGAATTTTACCACCGTGAAGTCAGCAA 183
Db 339 uProIleGlyLysArgGlyGlnIleSerGlySerPheSerValGlnGlyAlaAsnAs 359
Qy 184 ACTCGCCTCAGATTAAATCTGGAGCGATGCTCTTTGTCGCCAGGTTCTCAGTGAAGA 243
Db 359 pLeuAlaValLeuLeuArgAlaGlyAlaLeuProAlaThrLeuThrValValGluAr 379
Qy 244 GACGATCTCTCTGATCTGGGAAAAACAATGTACACAAGGCATATCTCAGCATGCTG 303
Db 379 gThrValGlyProSerLeuGlyAsnAspSerIleThrAlaGlyLeuThrAlaSerAlaI 399
Qy 304 TGGCTTGGCAATGCTATTGTTTGTAGCAGGATATATATAGATTGGAGGCGTCATCGC 363
Db 399 eGlyAlaValGlyValLeuIlePheMetPheValPheThrGlyPheGlyLeuLeuAl 419
Qy 364 TTCGGGAGCTCTTCTCTGAATCTTTTCTGCTATCTGGGCAGCTCTACAGTATTGGATGC 423
Db 419 aAsnIleAlaLeuIleValAsnValValMetLeuIleAlaValLeuSerValIleGlySe 439
Qy 424 GCCACTACCTTGTGACGACTCGCTGGGATGTTCTTCTGCTATGGGATGGCGGTAGATGC 483
Db 439 rThrLeuThrLeuProGlyIleAlaGlyIleValLeuThrIleGlyMetAlaValAepSe 459
Qy 484 AAATGTTCTCTATTGGAAGAATCCGAGAGGAATTTTATGTCTCAAGCTCTTAAAAA 543
Db 459 rAsnValLeuIleThrGluArgGileArgGluGluValLysSerGlyLysProLeuIleGl 479
Qy 544 ATCTGTAGAAAAAGGATATACCAAGCTTTTGGAGCCATTTTGTATTCTAATCTGACTAC 603
Db 479 nAlaLeuAspAsnGlyPheThrArgAlaPheAlaThrIleIleAsnLeuLeuThr 499
Qy 604 AGTATTGGCCTCAGCACTCTTTTCTCTAGATACAGGCGCTATTAAAGGGTTTGCTTT 663

Db 499 rLeuIleValAlaSerValLeuPheThrMetGlyThrGlyProValLysGlyPheAlaVa 519
Qy 664 GACATGATTATTAGGAATTTCTCTCAATGTTTACGGCTCTTTTCATGACATAAATTTT 723
Db 519 lThrLeuAlaValGlyIleIleThrThrValPheThrAlaThrThrLeuThrAlaTrpMe 539
Qy 724 CTTTCATCTCTGGATGAATTAAGACCCCAACATACACAGTTGCAAT-----ATGATGAA 774
Db 539 tPheGlyTyTrpValArgArgSerArgProLysHisLeuProLysGlyValArgThrAl 559
Qy 775 TAAGTTCTGGGATTAAGCATGATTTCTTGAGAGGATGCAAAAAAATTTGGGCTGTTTC 834
Db 559 aMetPheAspGlyArgAspIleProPheMetArgTyArgArgValValPheMetIleTh 579
Qy 835 TGAAGTGTGTTTCTTTTAGTGTGCTCTCGGTTTGGAGCTTGGGAATTCGCTTTT 894
Db 579 rGly---ValIleMetLeuAlaCysValGly---GlyPheValAlaLysGlyLeuAsnLe 597
Qy 895 GGGAAATCGATTTTAAAGAGGGGTATGCTTTTACCTTTAAATCCAAAGAGCATGCGCATCAG 954
Db 597 uGlyIleAspPheGlnGlyGlySerValIleGluValArgAlaLysGln---GlyGluAl 616
Qy 955 CGATGTTGCTCAATGCTGCGCAAGTTGTGCATAACTACAGGAAGCTGCTCTTTCTTC 1014
Db 616 aAspLeuAlaAspIleArg-----GluArgLeuAsnGlnLeuAsnLeuGly-- 631
Qy 1015 TAGAGACTTCCGTATTCAACATTTTGGATCTTCAGAAAAAGATCAAAATCTATTATTAGTGA 1074
Db 632 ----GluIleGlnAlaGlnAsnPheGlySerProGlnAspValLeuIleArgIle----- 648
Qy 1075 TAAAGCTTTAAGCTATATAAGCAGATACGAGCCTCTCTCTTAAATTTAAAGATCATCAG 1134
Db 649 -GlnAlaGlnAspGlyGlyGluAsnAlaGluGlnSerAlaIleThrLeu----- 664
Qy 1135 CTGGCGTATTGTCGGGATGTTGTGCAGAAACAGCCCTAGATTTCTCTACGGAACACTCAA 1194
Db 665 -----ValArgGlyGluLeu-----GluAspGly 672
Qy 1195 ACGAAACGCAAAATTTTGGTCAAAAGTAAGCAGCAAACTATCGAAGAAAATGCGTTATCA 1254
Db 672 sTyArgPheArgArgValGluValValGlyProAlaValSerGlyAspLeuThrValTh 692
Qy 1255 GGCACCATCGGCTTTTAGGAGCTTTGGCAATCATCTGCTCTATGATGATTTTGGCGTT 1314
Db 692 rSerThrIleGlyValLeuLeuAlaMetAlaIleMetValTyIleTrpValArgPh 712
Qy 1315 TGAATGCAATATGCTTTCAGTCCGCTATGCGCTTAAATTCATGACCTTTTGGCTACCTG 1374
Db 712 eGluTrpGlnPheAlaLeuGlyAlaValIleSerMetValHisAspValPheThr-- 731
Qy 1375 TGCAGTCTGTTTATAGCACATTTCTTTTCAAGAAAATTCAAATAGATTTTGAAGCAT 1434
Db 732 -----IleGlyLeuPheValPheLeuGlyIleGluPheAsnLeuThrSerIl 747
Qy 1435 TGGTGTCTTAATGACTATTGGGATTTTCAATTAACAATCTTTGATCATTTTGTATGTCG 1494
Db 747 eAlaAlaIleLeuThrIleIleGlyTySerLeuAsnAspThrValValIleTyAspAr 767
Qy 1495 TATTCTGTAAGATCGCAAGCAACCTGTTTACCCCTATGATGCTTTTGTAGTTAATGATGC 1554
Db 767 gValArgGluAsnLeuArgTyTyLysMetProLeuSerMetIleIleAspValSe 787
Qy 1555 CCCTCAAAAGCGCTTACGCGCAGCGTAAGTACACACAGCTACACTCTATCAGTTTGTGT 1614
Db 787 rLeuAsnGlnThrLeuSerArgThrIleLeuThrGlyLeuThrValLeuLeuAlaLe 807
Qy 1615 AATGCTTTGTTTATAGCGCGCTCTCTGCTTTTAAATTTTGCATTTTATATGACCATAGG 1674
Db 807 uAlaLeuTyLeuPheGlyGlyGluValIleArgSerPheThrPheAlaMetLeuPheGl 827
Qy 1675 GATTCTTCTAGGAATTTATGCTCTCTTTATATATGACCACTCTCTGTGTGTTG---TTTAT 1731

Db 827 yValGlyIleGlyValPheSerSerValTyrIleAlaIleProValIleuIleAlaPheIly 847
 QY 1732 GGTCCGTAAGAAATATCGC 1750
 Db 847 sleuAArgProAspSerIlys 853

RESULT 7

AB3337
 protein translocase chain secd / protein translocase chain secf [imported] - Brucella me

C/Species: Brucella melitensis
 C/Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 01-Feb-2002

C/Accession: AB3337

R.; DelVecchio, V.G.; Kapral, V.; Redkar, R.J.; Patra, G.; Mijer, C.; Los, T.; Ivanova,
 .; Mazur, M.; Goldsman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letess
 Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002

A/Title: The genome sequence of the facultative intracellular pathogen Brucella melitensis

A/Reference number: AD3252; PMID:11756688

A/Accession: AB3337

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-845 <KUR>

A/Cross-references: GB:AB008917; PID:AL51861.1; PID:G17982610; GSPDB:GN00190

A/Experimental source: strain 16M

C/Genetics:

A/Gene: BME10680

A/Map position: 1

Alignment Scores:

Score: 1.88e-54 Length: 845
 Percent Similarity: 52.87% Matches: 175
 Best Local Similarity: 30.43% Conservative: 129
 Query Match: 21.60% Mismatches: 198
 DB: 2 Gaps: 14

US-09-868-987-1 (1-1864) x AB3337 (1-845)

QY 80 GCTGTAGTATTGATGCGTTATATGTCACGACGCTTATTTAAACGTCCTCCATTGAAA--- 136
 Db 298 AAlIleValIleuAspGlyLysValIleuSerAlaProValIleArgGluProIleThrGly 317
 QY 137 AATCATGCGCATGCTCTCGGAAATTTACCGACGCGTGAAGAGCAAACTCGCCACATG 196
 Db 318 GlySerGlyGlnIleSerGlySerPheThrValGlnAspThrValIleuSerAlaIleu 337
 QY 197 TTAATAATCGAGCGATGCTTTGTTCCGAGGTTCTCACTGAGAGACGATCTTCT 256
 Db 338 LeuArgAlaGlyAlaIleuProAlaProIleuThrValIleGluGluArgThrValGlyPro 357
 QY 257 GATCTGGGAAAAAACAATGTACACAGGCAATTCATCAGCATGCTGCGCTGGCAATG 316
 Db 358 AspIleuGlyGlyAspAlaIleuSerMetGlyLeuMetThrGlyIleIleGlyPheIleu 377
 QY 317 CTATATGTTTGTATGAGGATATATATGATTTGAGGCGTCATGCTCGGAGCTGT 376
 Db 378 ValAlaValPheIleLeuIleuLeuIleuIleuIleuIleuIleuIleuIleuIleu 397
 QY 377 CTTCGATCTTTGCTTATCTATGCGGACCTCTACAGTATTTGAGTCCGCACTGCTTG 436
 Db 398 LeuLeuHISThrLeuIleuThrPheSerAlaIleuSerIleuIleGlyAlaThrLeuThrIleu 417
 QY 437 TCAGAGCTCGTGGATGTTCTTCTGTAATGGAGTGGCCGCTGATGCAATGTTCTTGA 496
 Db 418 ProGlyIleAlaGlyIleIleIleuGlyIleGlyIleAlaValAlaAspAlaAsnIleIleu 437
 QY 497 TTCGAAAGAAATCCGAGAGATTTTATATGCTCAAGCTTAAATAATCGTAGAAAAA 556
 Db 438 AsnGluArgIleArgGluGluThrArgIleGlyLeuGlyAlaMetAlaIleuAspIys 457
 QY 557 GGATATACCAAGCTTTTGGAGCACTTTTGTCTTAACTGATGACTGATTTGGCCTCA 616
 Db 458 GlyPheHISerAlaPheAlaThrIleValAlaAspAlaAsnValThrThrLeuThrAlaThr 477

QY 617 GCACTTCTTTCTTCCATAGTACAGGCGCTATTAAAGGTTGCTTGAACATGATTTA 676
 Db 478 IleLeuLeuPheLeuPheGlyThrGlyProValArgGlyPheAlaValIleuMetLeu 497
 QY 677 GGAATTTTCTTCAATGTTTACGCGCTTTTTCATGACATTAATTTTCTTCACTGCTGG 736
 Db 498 GlyIleAlaIleSerMetPheThrAspValThrLeuValArg---MetIleMetAlaTyr 516
 QY 737 ATGATATAG-----ACC 748
 Db 517 PheValArgArgArgIleuIleuValIleGluProPheLeuIysPheValPro 536
 QY 749 CAACATACAGATGTCATATGATGAAT---AGCTGCGGAGATTAAGATGATTTCTTG 805
 Db 537 GluHISThrAsnPheArgPheMetAsnAlaArgPheIleGlyIle----- 551
 QY 806 AGAGATGCAAAAAAATCTTGGCTGTTCTGGAAGTGTCTTTTGAAGTTGCGTTGCT 865
 Db 552 ---GlyValSerIleValIleuSerIleAla---SerIleIleuPheIysProGly 569
 QY 866 CTGGGCTTGGAGCTGGAATTCGCTTTTGGATGATTTTAAAGAGGCTATGCTTT 925
 Db 570 LeuAsnTyr-----GlyIleAspPheIysGlyIleGlnAla 582
 QY 926 ACCTTATATCCAAAGAGCATGCGATCAGCATGTTGCTCAAAATGCGTGCAGATTGTG 985
 Db 583 GluIleSerThrSerGlnPro-----AlaAspLeuAlaGlnIleuArgAlaIysLeu-- 599
 QY 986 CATTAACATACAGAGAGCTGCTTCTTCTTACAGACTTCGATTCAAATGATTTGATCT 1045
 Db 600 -----GlyAlaLeuAsnLeuGlyGluValAlaIleuGlnThrAlaGlySer 614
 QY 1046 TCAGAAAAAGTCAAATCTATTTT-----AGTATTAAGCTTTTAAAC 1087
 Db 615 ProAsnGlnValIleuIleArgValGlnArgGlnGluGlyGluGlnAlaGlnThrAla 634
 QY 1088 TATATACAGAGATAGAGCGCTCTCTCAAAATTAAGATCAATGAGCTGCGCTATGT 1147
 Db 635 AAlIleAsnIysMetArgGluAlaValThrGluLeuAsp-----Pro 648
 QY 1148 GGGATTTGTTTCAGAAAACAGGCTAGATTTCTACGGAACCTTAAACGAAACGCAAA 1207
 Db 649 GlyValIysIle-----GluArgThrGluVal 657
 QY 1208 TTTTGTCAAAGGTACAGCAAACTATGCAAGAAAATGCGTTATCAGGCGACATGGG 1267
 Db 658 ValGlyProIysValSerGlyGluLeuAlaArg-----SerGlyMetIleAla 673
 QY 1268 CTTTATAGAGCTTTGGCAATCATCTTGCCTATGATGAGTTGGCGCTTGAATGGCAATAT 1327
 Db 674 ValIleuSerAlaValAlaMetLeuPheThrIleThrPheArgPheGluThrPhePhe 693
 QY 1328 GCTTTCAGTCCCGTATGCGCTTAAATTCATGACCTTTTGGCTACCTGTGACGCTTGT 1387
 Db 694 AlaIleuGlyAlaIleAlaThrIleuIleuAsp-----ThrThrIysIle 708
 QY 1388 ATACACATTTCTTTTGAAGAAAATTCAAATATGATTTGCAAGCATGCTGCTTTATG 1447
 Db 709 IleGlyPhePheAlaIleuThrGlnIleuAspPheAsnIleuThrAlaIleAlaIleu 728
 QY 1448 ACTGATTTGGGATATTCATTAACAATACCTTATGATCTTTTGAATGCGTATTCGAGAT 1507
 Db 729 ThrIleIleGlyIysSerValAsnAspIysValValIysArgPheMetArgGluAsn 748
 QY 1508 CGCAAGCGAAACCTGTTTACCCCTATGACATGTTTATGATATGATGCGCTCAAAAGACG 1567
 Db 749 MetArgLeuIysIysSerIysThrIleuArgGluIleIleAspMetSerIleAsnGlnVal 768
 QY 1568 TTCAGCCGACGGTATACACACAGCTCAATCTATCACTTGTGTTATGCTTTGCTT 1627
 Db 769 LeuValArgCysIleIysThrSerMetThrThrPheLeuIysMetPhePrometAlaIle 788
 QY 1628 ATAGCGGCTTCCTGCTTATTTGATTTATATATGACATGAGGATTTCTTCTGGA 1687

Db 789 TrpGlySerAlaValHisAsnPheAlaValProMetLeuPheGlyValIleAla 808
 Qy 1688 ACTTATCGTCTTTATATTGACCACTCTGTGTGTTTATG 1732
 Db 809 ThrSerSerIlePheIleAlaAlaProIleLeuLeuLeu 823
 RESULT 8
 D89946
 protein-export membrane protein SecDF [imported] - Staphylococcus aureus (strain N315)
 C:Species: Staphylococcus aureus
 C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
 C:Accession: D89946
 R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Ogino,
 ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;
 C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
 Lancet 357, 1225-1240, 2001
 A:Title: Whole genome sequencing of methicillin-resistant *Staphylococcus aureus*.
 A:Reference number: A89758; MUID:21311952; PMID:11418146
 A:Accession: D89946
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-759 <KUR>
 A:CROSS-references: GB:BA000018; PID:g13701435; PIDN:BAB42729.1; GSPDB:GN00149
 A:Experimental source: strain N315
 C:Genetics:
 A:Gene: secF

Alignment Scores:
 Pred. No.: 8.8e-52 Length: 759
 Score: 685.00 Matches: 171
 Percent Similarity: 53.01% Conservative: 137
 Best Local Similarity: 29.43% Mismatches: 197
 Query Match: 20.68% Indels: 76
 DB: 2 Gaps: 14

US-09-868-987-1 (1-1864) x D89946 (1-759)

Qy 104 CTCAGCAGCCCTATTAAACGTCCTCAATGAAATCATGCC---AGTGTCTCAGGGAAA 160
 Db 197 IleSerAlaAlaSerValAspGlnProIleAsnSerAspSerValGluIleSerGly 216
 Qy 161 TTTACCCACCGTGA-----GTGAGCAACTCGCTCAGATCTCTCTGATCTTGGGAAA 268
 Db 217 PheLysGlyGlnGluGlyValLysAlaLysGlnIleAlaGluLeuLeuAsnAlaGly 236
 Qy 209 CGCATGCTTTTGTTCGCGAGTCTCAGTCAAGAGAGATCTCTCTGATCTTGGGAAA 268
 Db 237 SerLeuProValAspLeuLysGluIleTyrSerAsnSerValGlyAlaGlnPheGlyGln 256
 Qy 269 AAACAATGTACACAGGCATTATCTCAGCATGCTGTGGCTTGGCAATGCTTATTGTTTG 328
 Db 257 AspAlaLeuAspLysThrValPheAlaSerPheIleGlyValAlaLeuLysLeuPhe 276
 Qy 329 ATGACGTATATTAGATTGGAGGCTCATCGTTCGGAGCTGTCTCTGAATCTT 388
 Db 277 MetLeuGlyPheTyrArgLeuProGlyLeuValAlaIleAlaLeuThrThrTyrIle 296
 Qy 389 TTGCTTATCTGGCAGCTCTACAGTATTGTGATGGCCACTCACCTCTCAGGACTCGCT 448
 Db 297 TyrLeuThrLeuValAlaPheAsnPheIleSerGlyValLeuThrLeuProGlyLeuAla 316
 Qy 449 GGGATTGTCTGTATGGGATGGCGGTAGATGCAATGTCTTGTATTTCGAAGAATC 508
 Db 317 AlaLeuValLeuGlyValGlyMetAlaValAspAlaAsnIleIleMetTyrGluArgIle 336
 Qy 509 CGAGAGGAATTTTATTGCTCTCAAGTCTTAAATCTGTAGAAAAGGATATACCAAG 568
 Db 337 LysAspGluLeuArgIleGlyArgThrIleLysGlnAlaPheSerLysAlaAsnLysSer 356
 Qy 569 GCTTTTGACCATTTTGTATTGAATCTTAACTTACAGTATTGGCTCAGCACTCTTTTC 628
 Db 357 SerPheLeuThrIlePheAspSerAsnLeuThrThrValIleAlaAlaValLeuPhe 376

Qy 629 TTCTAGATACAGGCGCTATTAAAGGCTTTGCTTTGACATTGATTTTAGGAATTTCTCT 688
 Db 377 PhePheGlyGluSerSerValLysGlyPheAlaThrMetLeuLeuLeuGlyIleLeuMet 396
 Qy 689 TCAATGTTTACGGCTCTTTTCATGACTAAATTTTCTTCATGCTG----- 733
 Db 397 IlePheValThrAlaValPheLeuSerArgPheLeuLeuSerLeuValSerSerAsn 416
 Qy 734 -----TGGATGAATAAGACCCCAACATACACATGTCATATGATGAAT 775
 Db 417 IlePheLysAsnGlnPheTrpLeuPheGlyValLysLysAsnLysArgHisAspIleAsn 436
 Qy 776 AAGTTCCTGGGATA-----AAGCATGATTTCTTCAGAGGA 811
 Db 437 GluGlyValAspValHisAspLeuLysThrSerPheGluLysTrpAsnPheValLysLeu 456
 Qy 812 TGCAAAAACATTTGGGCTGTTCTGGAAGTGTCTTTTCTTTAGTTCGCTCTCGG 871
 Db 457 AlaLysProLeuIleGlyValSerIleLeuIleValValGlyLeuValIleLeuTyr 476
 Qy 872 TTTGGAGCTCGAATTCGCTTTTGGGAATGGAATTTTAAAGAGGGGTATGCTTTTACCTTT 931
 Db 477 IlePheLysLeuAsn-----LeuGlyIleAspPheSerSerGlyThrArgValAspPhe 494
 Qy 932 AATCCAAAAGAGCATGCATCAGCGATGTTGCTCAAAATGCGTGGCAAAAGTTGTGCATAA 991
 Db 495 GlnSerLysGln-----AlaIleThrGlnGlnLysValGluGlnVal 508
 Qy 992 CTACAGGAAGCTGCTTTCTCTTAGAGACTTCGCTATTCAAAACATTTGGATCTTCAGAA 1051
 Db 509 ValLysAspSerGlyLeuLysAlaAsp-----GlnIleGlnIleAsnGlyLysAspAsn 526
 Qy 1052 AAGATC---AAAATCTATTATTAGTATAAAAGCTTTAAGCTATATAAGACAGATACGAGCC 1108
 Db 527 LysValAlaThrValGlnPheLysAsp----- 536
 Qy 1109 TCTCTCTAAATTAACGATCATGAGCTGGCGTTATTGTGGGATTGTTGTGAGAAACAGG 1168
 Db 537 -----LeuThrArgAlaGln 541
 Qy 1169 CCTAGATTTCTACGAAACTCTAAACGAAACGCAAAATTTGG----- 1213
 Db 542 AspAsnLysLeuSerAspAsnIleLys-----SerLysPheGlyAspThrProGlnIle 559
 Qy 1214 TCAAGGTAAAGCAGCAAACTATCGAAGAAATGCTTATCAGGCGACCATCGGCTTTTA 1273
 Db 560 AsnThrValSerProIleIleGlyGlnGluLeuAlaLysAsnAlaMetLeuAlaLeuIle 579
 Qy 1274 GGAGCTTTGGCAATCATCTTTGCTCTATGTCAGTTTGGCTTTGAATGCAATATGCTTTC 1333
 Db 580 TyrAlaSerIleGlyIleIleIleTyrValSerLeuArgPheGluTrpArgMetGlyLeu 599
 Qy 1334 AGTCCGTATGCGCTTTAAATTCATGACCTTTTGGCTACCTGTGTCAGCTCTGTTTATAGCA 1393
 Db 600 SerSerValLeuAlaLeuLeuLeuHisAspValPhe-----IleIleValAla 614
 Qy 1394 CATTTCTTTTGAAGAAATTCAAATAGATTTCGAGCCCATTTGCTTTAATGACCTGTA 1453
 Db 615 IlePheSerLeuPheArgIleGluValAspLeuThrPheIleAlaAlaValLeuThrIle 634
 Qy 1454 TTGGGTATTTCATTAAACAATCTTTGATCATTTTTCATCGTATTCGTGAAGATCGCAA 1513
 Db 635 ValGlyTyrSerIleAsnAspThrIleValThrPheAspArgValArgGluAsnLeuGln 654
 Qy 1514 ---CGCAACCTGTTTACC-----CCTATGATGTTTTTGTATGATGCTCCCTCAA 1561
 Db 655 LysValLysValIleThrThrThrGluGlnIleAspAspIleValAsnArgSerIleArg 674
 Qy 1562 AAGACGTTACGCGCAGGTAATGACAAACAGCTACACTCTATCAGTCTTGTGTTAATGCTT 1621
 Db 675 GlnThrMetThrArgSerIleAsnThrValLeuThrValIleValValValAlaIle 694

QY 1622 TTGTTAAGGCGCTCCTCTTCTTAAATTTGATTTATGATTCATGACATGAGGATCTT 1681
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
Db 695 LeuphepeglyalaProthrllephensnPhetrleuAlaleuPheliledyleuLe 714
QY 1682 CTAGGAATTTATCGTCTCTTCTTAAATTTGACACCACTCTGTTGTTGTTAAAGTCCGTA 1741
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
Db 715 SerglyValpheserSerlelePhelileAlaValProleutPpGlyllemelyleysarg 734
QY 1742 GAA 1744
: : :
Db 735 Gln 735
RESULT 9
H69704
protein-export membrane protein secF - Bacillus subtilis
C/Species: Bacillus subtilis
C/Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 15-Oct-1999
C/Accession: H69704
R/Kunst, F.; Ogasawara, N.; Mooser, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter
C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Cho
A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997
A/Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizni, A.; Galler
lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.
Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
A/Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mueel
Y, M.; Ogawa, K.; Ogilwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Potteville
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon,
A/Authors: Schleich, S.; Schroeder, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Serot
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpetra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
A/Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.
A/Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A/Reference number: A69580; MUID:98044033; PMID:9384377
A/Accession: H69704
A/Status: preliminary; nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-571 <KUN>
A/Cross-references: GB:259118; GB:AL009126; NID:G2635200; PIDN:CAB14724.1; PID:el184013;
A/Experimental source: strain 168
C/Genetics:
A/Gene: secF
Alignment Scores:
Pred. No.: 3.1e-51 Length: 571
Score: 678.50 Matches: 169
Percent Similarity: 50.43% Conservative: 125
Best local Similarity: 28.99% Mismatches: 192
Query Match: 20.49% Indels: 97
DB: Gaps: 15
US-09-868-987-1 (1-1864) x H69704 (1-571)
QY 104 GTCAGACCCCTATTTAAAGCTCCATGAA--AATCATGCCAGTGTCCAGGAAA 160
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
Db 31 ValSerAlaProbenValSerGlnIleuasnThrThrAspValLysIleGluIyis 50
QY 161 TTTACCCACCCGAGTAGAGCAAACTCGCTCAGATTTAAATTTGAGCGATGCTTTT 220
| | | | | : : : | | | | | : : : | | | | | : : : | | | | | : : :
Db 51 PheThrAlaGlnGlnAlaLysAspLeuAlaSerIleLeuAsnAlaGlyAlaLeu----- 68
QY 221 GTTCCCGAGGTTCTCAGTAAGAG-----ACGATCTCTTCGATCTTGGGAAAAA 271
| | | | | : : : | | | | | : : : | | | | | : : : | | | | | : : :
Db 69 --ProValLysLeuThrGlnLysSerThrSerValGlyAlaGlnPhedGlyGlnGln 87
QY 272 CAATGTACCAAGGATATTCACGATGCTGCTTGGCGAAGCTTATGTTTGAAG 331
| | | | | : : : | | | | | : : : | | | | | : : : | | | | | : : :
Db 88 AlaLeuHisAspThrValAlaPheAlaGlyLeValGlyIleAlaIleIlePheLeuPheMet 107
QY 332 AGCGTATATTTAGATTTTGGAGCGCTCATCGCTTGGAGGCTGTTCTTGAATCTTTG 391
| | | | | : : : | | | | | : : : | | | | | : : : | | | | | : : :
Db 108 LeuPheTyrrTyrrArgLeuProGlyLeuIle-----AlaValIleThrLeuSerVal 124

QY 392 CTATATCGGACAGCTACAG-----TATTTGGATGCGCACTCACCTTGTCAGA 442
| | | | | : : : | | | | | : : : | | | | | : : : | | | | | : : :
Db 125 TyrIleTyrrIleThrLeuGlnIlePheAspTrpMetCsmAlaValLeuThrLeuProGly 144
QY 443 CTCGCTGGAGTGTCTTCTGATGAGGATGCGGTAGATCAAAATGTTCTTGTATGAA 502
| | | | | : : : | | | | | : : : | | | | | : : : | | | | | : : :
Db 145 IleAlaAlaLeuIleLeuGlyValGlyMetAlaValAspAlaAsnIleIleThrTyrrGlu 164
QY 503 AGAATCCGAGAGAAATTTTATTTGTCTCAAAAGCTTAAAAATCTGTGAAAAAGATAT 562
| | | | | : : : | | | | | : : : | | | | | : : : | | | | | : : :
Db 165 ArgIleLysGlnGlnLeuLysSerValArgSerAlaPheArgSerGlyAsn 184
QY 563 ACCAAGGCTTTGGAGCATTTTGAATTCATGCTACGATGCTTGGCTGAGCACT 622
| | | | | : : : | | | | | : : : | | | | | : : : | | | | | : : :
Db 185 ArgArgSerPheAlaThrIlePheAspAlaAsnIleThrThrIleIleAlaValAl 204
QY 623 CTTTCTTCTCTAGTACAGGCGCTATTTAAAGGTTTGGCTTGTGATTTTGAAT 682
| | | | | : : : | | | | | : : : | | | | | : : : | | | | | : : :
Db 205 LeuPheIlePheGlyThrSerSerValLysGlyPheAlaThrMetLeuIleLeuSerIle 224
QY 683 TTCTCTCAATGTTTACGCTCTTTTCACTAATAATTTTCTTCATGCTG----- 733
| | | | | : : : | | | | | : : : | | | | | : : : | | | | | : : :
Db 225 LeuThrSerPheIleThrAlaValPheLeuSerArgPheLeuLeuAlaLeuValGlu 244
QY 734 -----TGATGAAATAG----- 745
| | | | | : : : | | | | | : : : | | | | | : : : | | | | | : : :
Db 245 SerArgTrpLeuAspArgLysGlyTrpPheGlyValAsnLysLysIleMetAsp 264
QY 746 -----ACCAACATACAGAGTTGCATATGATGAAATAGTTCGAGGATAG 793
| | | | | : : : | | | | | : : : | | | | | : : : | | | | | : : :
Db 265 IleGlnAspThrAspGlnAsnThrGlnProHisIleThrProGln-----Lys 280
QY 794 CATGATTTCTTGAGAGATGCAAAAAATTGGGCTGTTCTGGAAGTGTTTTCTTTA 853
| | | | | : : : | | | | | : : : | | | | | : : : | | | | | : : :
Db 281 TrpAspPheThrSerLysArgLysTyrrPhePheIlePheSerSerAlaValAla 300
QY 854 GGTGCGTCTCTCGGCTTGGAGCTCGAATTCGTTTGGGAATGATTTTAAAGA 913
| | | | | : : : | | | | | : : : | | | | | : : : | | | | | : : :
Db 301 GlyIleIleIleLeuLeuValPheArgLeuAsn-----LeuGlyLysPheAlaSer 318
QY 914 GGG-----TATGCTTTACCTTTTATCCAAAAGCATGCG 949
| | | | | : : : | | | | | : : : | | | | | : : : | | | | | : : :
Db 319 GlyAlaArgIleGlnValGlnSerAspHisLysLeuThrThrGlnGlnValGlyLysAsp 338
QY 950 ATCAGCATGTTGCTCAATATGCGTGCAGAACTTTGTCATTAATACAGAGAGCTGCTT 1009
| | | | | : : : | | | | | : : : | | | | | : : : | | | | | : : :
Db 339 PheGluSerLeuGlyMetAspProAspThrValValLeuSerGlyLysSerAsnIle 358
QY 1010 TCT---TCTAGAGACTTCGTATTTCAAAACATTTGGATCTTCAGAAAAGATCAAAATCTAT 1066
| | | | | : : : | | | | | : : : | | | | | : : : | | | | | : : :
Db 359 GlyValAlaArgPheValGlyValProAspLysGlnThrIleAlaLysValLysThrTyrr 378
QY 1067 TTTAGTATTAAGCTTTAAGCTATTAAGCAGATACGACCTCTCTTAAATTAACG 1126
| | | | | : : : | | | | | : : : | | | | | : : : | | | | | : : :
Db 379 PheLysAspLys----- 382
QY 1127 ATCATGAGCTGGGTTATTTGTGGATTTGTTGCAAAACAGGCTTACATTTCTTACGA 1186
| | | | | : : : | | | | | : : : | | | | | : : : | | | | | : : :
Db 383 -----TYGly 384
QY 1187 AACTCTAAGAAACGAAATTTTGTGCAAGGTACAGCAAGCACTATGAGAAATG 1246
| | | | | : : : | | | | | : : : | | | | | : : : | | | | | : : :
Db 385 SerAspProbenValSerThrValSerProThrValGlyLysGlnLeuAlaArgAsnAla 404
QY 1247 CGTATCAGCGCAACATCGGCTTTTGAAGCTTTGGCAATCATCTTGTATGTAGT 1306
| | | | | : : : | | | | | : : : | | | | | : : : | | | | | : : :
Db 405 LeuTyrrAlaValAlaIleAlaSerIleGly-----IleIleIleTyrrValSer 420
QY 1307 TTGCGCTTTGAAGGCAATAGCTTTTCAAGTCCGCTATGCTTATTCATGACCTTTTG 1366
| | | | | : : : | | | | | : : : | | | | | : : : | | | | | : : :
Db 421 IleArgPheGlyLysTyrrLysMetAlaIleAlaIleAlaIleAlaSerLeuLeuTyrrAsp----- 438
QY 1367 GCTACCTGTGACGCTTGTGTTATAGCATTCTTTTGAAGAAATTCAAATGATTTG 1426

```
Db 439 -----AlaPhePheIleValThrPhePheSerIleThrArgLeuGluValAspVal 455
Qy 1427 CAAGCCATTGGTCTTTAATGACTGTATTGGGGTATTCTAATAAACAATACATTTTGATCAT 1486
Db 456 ThrPheIleAlaAlaIleLeuThrIleIleGlyTyrSerIleAenAspThrIleValThr 475
Qy 1487 TTTGATCGTATTCTGTAA-----GATCGCAAGCGAACCTGTTTACCCCTATGCAAT 1537
Db 476 PheAspArgValArgGluHisMetLysLysArgLysProLysThrPheAlaAspLeuAen 495
Qy 1538 GTTTTAGTAAATGATGCCCTTCAAAAGACGTTTCAGCGCACGGTAATGACACAGCTACA 1597
Db 496 HisIleValAenLeuSerLeuGlnGlnThrPheThrArgSerIleAenThrValIleThr 515
Qy 1598 ACTCTATCAGTCTTTTAAATGCTTTTGTATAGCGCGCTCTCTGCTCTTAAATTTTGA 1657
Db 516 ValValIleValValThrLeuLeuIlePheGlyAlaSerSerIleThrAsnPheSer 535
Qy 1658 TTTATATGACCATAGGATCTCTTAGGAATCTTTATCGTCTCTTTATATATGACCACT 1717
Db 536 IleAlaLeuLeuValGlyLeuLeuThrGlyValTyrSerSerLeuTyrIleAlaAlaGln 555
Qy 1718 CTGTTGTTG 1726
Db 556 IleTrpLeu 558

RESULT 10
F75350
C:probable protein-export membrane protein - Deinococcus radiodurans (strain R1)
C:Species: Deinococcus radiodurans
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Mar-2000
R:White, O.; Eissen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
S.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A:Reference number: A75250; MUID:20036896; PMID:10567266
A:Accession: F75350
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-768 <WHI>
A:Cross-references: GB:AE002022; GB:AE000513; NID:g6459590; PIDN:AAF11371.1; PID:g645959
A:Experimental source: strain R1
C:Genetics:
A:Gene: DR1822
A:Map position: 1

Alignment Scores:
Pred. No.: 2,48e-50 Length: 768
Score: 668.50 Matches: 181
Percent Similarity: 51.22% Conservative: 113
Best Local Similarity: 31.53% Mismatches: 223
Query Match: 20.18% Indels: 57
DB: 2 Gaps: 13

US-09-868-987-1 (1-1864) x F75350 (1-768)

Qy 77 ATGGCTGTAGTATTGATGCGGTATATGTCAGCAGCCCTATTTTAAAGTCCTCCATTGAAA 136
Db 230 MetAlaValValLeuAspAspGlnIleGlnSerValAlaThrIleAenGlnArgLeuPhe 249
Qy 137 AATCATGCGCAGTGTCTCAGGGAATTTTACCACCGTGAAGTGACCAACTCGCCTCAGAT 196
Db 250 ArgAspIleGlnIleSerGlyAsnPheThrProGluAlaSerGlnLeuVal 269
Qy 197 TTAATACTGAGCGATGCTCTTTGTTCCCGAGGTCTCTCAGTGAAGACGATCTCTCT 256
Db 270 LeuLysSerGlyAlaLeuProIleLysIleValThrAlaAlaGluArgSerIleGlyPro 289
Qy 257 GATCTTGGAAAAACAATGTACACAGCAATTATCTCAGCATGCTGTGGCTTGGCAATG 316
Db 613 TyrValGlyPheArgPheAspPheIleMetGlyLeuGlySerIleIleAlaAlaIleHis 632
```

```
Db 290 SerLeuGlyAlaAspAlaIleArgSerGlyAlaIleAlaLeuValGlyIleGlyLeu 309
Qy 317 CTTATTCTTTTGTAGTACGCTATATTAT---AGATTTGGAGCGCTCATCGCTTCGGAGCT 373
Db 310 ValPheValMetLeuPheAlaTyrTyrGlyLeuTrpPheGlyLeuValGlyAlaLeuGly 329
Qy 374 GTTCTTCTGAATCTTTTGGCTTATCTGGCGAGCTCTACAGTATTTGGATGCGCCACTCACC 433
Db 330 LeuLeuPheSerSerIleIleIleLeuGlyIleLeuGlyGlyPheGlyAlaThrLeuThr 349
Qy 434 TTGTCACGAGCTCGCTGGGATTTGTTCTTCTGATGGGATGCGCGTAGATGCAAAATGTTCTT 493
Db 350 LeuProGlyIleAlaGlyLeuValLeuThrIleGlyAlaAlaAspGlyAsnValIle 369
Qy 494 GTATTGAAAGAAATCCGAGAGGAATTTTATTGCTCAAAAGCTCTTAAATAATCTGTAGAA 553
Db 370 SerPheGluArgIleLysGluLeuAlaArgGlyLysGlyIleLysAenAlaIleGly 389
Qy 554 AAGGATATACCAAGGCTTTTGGAGCCATTTTGTATCTTAACCTTGATCTACAGTATTGGCC 613
Db 390 AlaGlyTyrGluHisSerThrAlaAlaIleLeuAspValAenAlaSerHisLeuLeuSer 409
Qy 614 TCAGCACTCTCTTCTTCTTAGATACAGGCGCTTATAAGGGTTTGTCTTGACATTGATT 673
Db 410 AlaLeuAlaLeuTyrAenTyrSerThrGlyAlaValLysGlyPheAlaValThrLeuIle 429
Qy 674 TTAGGAATTTTCTTCAATGTTTACCGCTCTTTTCATGACTTAATTTTCTTCATGCTG 733
Db 430 IleGlyValIleAlaSerThrPheSerAsnLeuValPheAlaLys---TrpPheMetGln 448
Qy 734 TGGATGAATTAAGCCCAACATACACAGTTCGATATGATGAATAAGTTCGTGGGATAAAG 793
Db 449 TrpLeuAlaGlnArgArgProAsn-----MetSerAlaProGlnTrpIleLys 464
Qy 794 CNT-----GATTTCTTGAGAGGATGCAAAAACCTTTGGGCTGTTCTGGAAGTGT 844
Db 465 HisThrHisPheAspPheMetLysProAlaLysValIleThrThrLeuSerValLeuLeu 484
Qy 845 TTTCTTTTAGTTCGCTGCTCGGGTTCGGGCTTGGAGCTTGGAAATCCGTTTGGGAATGAT 904
Db 485 AlaLeuAlaGlyAlaAlaLeuValAlaThrArgGlyLeuAen-----TyrGlyValAsp 502
Qy 905 TTTAAAGAGGGTATGCTTTTACCTTTTAAATCCAAAGAGCATGCGCATCAGCGATGTGCT 964
Db 503 PheAlaProGlyThrThrLeuThrAlaArgValAspArgGlnValThrThrGlu----- 520
Qy 965 CAAATGCGTGCACAAAGTTGTGCATAAACTACAGGAAGCTGCTCTTCT----- 1012
Db 521 GlnLeuArgAenSerValIle-----GlyAlaGlyValSerLysValThrGly 536
Qy 1013 -----TCTAGAGACTTCCTGATTTCAACATTTTGGATCTTCAGAAAAGATC 1057
Db 537 GlnSerAlaThrIleGlnArgAspThrThrProGlyGlnGlnGlnAsnPheThrVal 556
Qy 1058 AAAATCTATTATTAAGTAAAGCTTTAAGCTATATAAGCAGATACGAGCTCTCTCCTA 1117
Db 557 LysValProGluLeuAenAspAla-----GluValLysGlnIleGlyAlaAlaIleGly 574
Qy 1118 AAATTAACGATCATGACTGCGGTATTGTTGGGATTTGTTGTCAGAAAACAGCCCTAGATT 1177
Db 575 LysLeuPro----- 577
Qy 1178 CTCTACGAAACTCTAAACGAAACGCAAAATTTTGGTCAAAAGTAAGCAGCAAACTATCG 1237
Db 578 ---GlnGlyGlnValLeuAlaSerGluThrValGlyProAlaValGlyLysGluLeuThr 596
Qy 1238 AAGAAAATGCGTTTATCAGCGCACCATCGGGCTTTTAGGAGCTTTTGGCAATCATCTTGCTC 1297
Db 597 GlnLysThrIleTyrAlaValLeuLeuGly-----LeuGlyLeuLeuLeuVal 612
Qy 1298 TAGTGTAGTTGCGCTTTGAATGGCAATATGCTTTTACGTCCGTATGCGCTTTTAAATCAT 1357
Db 613 TyrValGlyPheArgPheAspPheIleMetGlyLeuGlySerIleIleAlaAlaIleHis 632
```


Db 595 ArgAlaGlyValLeuAlaValIleLeuSerLeuIleGly-----IlePheIle 610
 Qy 1298 TATGTAGTTTCGCTTTGAATGCAATATGCTTTCAAGTCGCGATGCGCTTTAATTCAT 1357
 Db 611 TyrValTrpPheArgPheArgTrpGlnLeuAlaLeuGlyAlaValLeuSerThrLeuHis 630
 Qy 1358 GACCTTTGGCTACTGTGCGAGTCTTGTATTATAGCACATTTCTTTTGAAGAAATTCAA 1417
 Db 631 AspVal-----ValIleLeuSerGlyMetPheIleValPheArgMetGlu 645
 Qy 1418 ATAGATTTCGAAGCATGGTGTCTTAATGACTGTATTGGGTATTCATTAAACAATACT 1477
 Db 646 PheAsnLeuTrpSerValAlaAlaIleLeuThrIleIleGlyTyrSerLeuAsnAspThr 665
 Qy 1478 TTGATCATTTTGTATCGPATTCGTAAGATCGCCAAGCAACCTGTTTACCCCTATGCAT 1537
 Db 666 ValValIleTyrAspArgValArgGluAsnLeuArgArgTyrLysSerAlaProLeuPro 685
 Qy 1538 GTTTTAGTAAATGATGCCCTTCAAAAGCGTTACGCCGACGTTAATGACAACAGCTACA 1597
 Db 686 AlaIleIleAspAlaSerIleAsnGlnThrLeuSerArgThrLeuLeuThrSerPheVal 705
 Qy 1598 ACTCTATCAGTTTGTGTAATGCTTTGTTTATAGCGGCTCTCTGCTCTTAAATTTGCA 1657
 Db 706 ThrPheLeuAlaHisValProLeuTyrAlaPheGlyGlySerGluIleArgMetPheAla 725
 Qy 1658 TTTATTATGACCATAGGATTTCTTCAAGAACTTTATGCTCTCTTATATATGCAACCACT 1717
 Db 726 LeuAlaLeuSerValGlyIleIleValAlaSerTyrSerSerIlePheIleAlaAlaPro 745
 Qy 1718 CTGTTGTTG 1726
 Db 746 LeuLeuVal 748
 RESULT 12
 T34673
 probable SecDF protein-export membrane protein - Streptomyces coelicolor
 C;Species: Streptomyces coelicolor
 C;Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 05-Nov-1999
 C;Accession: T34673
 R;Saunders, D.C.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrall, B.G.; Rajandream, M.
 submitted to the EMBL Data Library, December 1998
 A;Reference number: Z21552
 A;Accession: T34673
 A;Status: preliminary; translated from GB/EMBL/DBJ
 A;Molecule type: DNA
 A;Residues: 1-795 <SAU>
 A;Cross-references: EMBL:AL034446; PIDN:CAA22394.1; GSPDB:GN00070; SCOEDB:SCIA9.24C
 A;Experimental source: strain A3(2)
 C;Genetics:
 A;Gene: SCOEDB:SCIA9.24C
 Alignment Scores:
 Pred. No.: 5,99e-47 Length: 795
 Score: 630.00 Matches: 181
 Percent Similarity: 48.17% Conservative: 109
 Best Local Similarity: 30.07% Mismatches: 212
 Query Match: 19.02% Indels: 100
 DB: 2 Gaps: 14
 US-09-868-987-1 (1-1864) x T34673 (1-795)
 Qy 74 CGTATGGCTAGTATGATGAGGTTATATGTGTCAGCAGCCCTATTTTAAACGTCCCATTTG 133
 Db 220 ArgValAlaIleValLeuAspLysIleIleSerProGlnValAspProSerVal 239
 Qy 134 AAA-----AATCATGCCAGTGTCTCAGGGAATTTTACCCACCGT 172
 Db 240 SerCysGlyAlaGlyIleThrGlyGlySerThrGlnIleThrGlySerPheAspAspAla 259
 Qy 173 GAAGTGACAAACCTCGCTCAGATTAAATCTGGAGGATGTCTTTTGTCCCGAGTT 232
 Db 260 GluAlaArgGluLeuAlaLeuLeuIleLysGlyAlaLeuProValProValGluThr 279

Qy 233 CTCAGTGAAGAGACGATCTCTTCTGATCTTTGGGAAAAACAATGTACACAAGGCATTATC 292
 Db 280 IleGluGlnArgThrIleGlyAlaThrLeuGlyAspGluAlaIleAspAlaGlyAlaTrp 299
 Qy 293 TCAGCATGCTGTGCTGGCAATGCTTATTGTTTGTATGAGCGTATATTATAGATTGGA 352
 Db 300 AlaAlaValIleGlyThrAlaLeuThrAlaLeuPheIleIleValValTyrArgLeuMet 319
 Qy 353 GGCGTCAATCCCTTCGGGAGCTGTTCTTCTGAATCTTTTGTCTTATCTGGCAGCTCTACAG 412
 Db 320 GlyAlaLeuAlaThrValAlaLeuLeuCysTyrGlyLeuIleSerTyrAlaAlaLeuAla 339
 Qy 413 TATTGGATCGGCACATCACCCTTGTCTAGGACTCGCTGGGATTGTTCTTGTATGGGATG 472
 Db 340 AlaValGlyAlaThrLeuThrLeuProGlyLeuAlaGlyPheValLeuAlaIleGlyMet 359
 Qy 473 GCCTGATGATCAATGCTTCTGTTATTCGAAGAATCCGAGAGGAA-----TTTTATTG 526
 Db 360 AlaValAspAlaAsnValLeuValPheGluArgAlaArgGluGluGlnAlaAlaArgThr 379
 Qy 527 TCTCAAGTCTTAAATAATCTGTAGAAAAAGGATATACCAAGGCTTTTGGAGCCATTTT 586
 Db 380 ArgProSerThrArgSerAlaLeuThrAlaGlyPheArgSerAlaPheSerAlaIleAla 399
 Qy 587 GATTCTAATCTGACTCAGTATTGGCTCAGCATTCTTTTCTCTAGTACAGGCGCT 646
 Db 400 AspSerAsnIleThrThrLeuIleAlaAlaLeuLeuPhePheLeuAlaSerGlyPro 419
 Qy 647 ATTAAGGGTTGCTTGTGACATGATTAGGAAATTTCTCTCAATGTTTACGGCTCTT 706
 Db 420 ValTyrGlyPheGlyValThrLeuGlyIleGlyValLeuAlaSerMetValSerAlaLeu 439
 Qy 707 TTCATGACTAAATTTTCTTCATGCTGTGATGAATAAGACCAACATACACATGTCAT 766
 Db 440 ValIleThrArg-----ValLeuAlaGluPheAlaAlaSerArgProAlaValPhe 456
 Qy 767 ATGATGAATAGTTTCGTGGGATA----- 790
 Db 457 ArgArgProArgIleThrGlyIleSerSerThrGlyProValArgAspAlaLeuLeuArg 476
 Qy 791 AAGCATGATTCTTGTAGAGAGATGCAAAAACTTTGGCTGTTCTTCTGGAAGTGTTCCTT 850
 Db 477 ArgAspProPheLeuMetArgArgProArgTrpLeuAlaAlaSerLeuIleValLeu 496
 Qy 851 TTA-----GGTTGGTGTCTCTCGGGTTTGGAGCTCGAATTCGTTTTCGGA 898
 Db 497 ValValAlaGlySerGlyIleLeuValArgGlyLeuAsn-----PheGly 511
 Qy 899 ATGGATTTTAAAGGAGGTATGCCCTTTACCTTTAATCCA----- 937
 Db 512 IleGluPheThrGlyGlyArgLeuIleGluTyrSerThrAlaThrGlnValAspProAsp 531
 Qy 938 ---AAAGAGCATGCATCAGCGATGTTGCTCAAAATGCTGGCAAAAGTTGTGCATAACTA 994
 Db 532 ArgAlaArgAspAlaLeuAlaAspAlaGlyPheProArgAlaValVal----- 547
 Qy 995 CAGGAAGCTGCTTCTTCTCTAGAGACTTCGCTATTCAACATTTGGATCTTCAGAAAAG 1054
 Db 548 GlnSerSerGlyAspGlyAspLeuThrValArgThrGluLeuLeuThrAspThrGluAla 567
 Qy 1055 ATCAAAATATTATTAGTATGAAGCTTTAAGC-----TATACCTAAAGCAGATA 1102
 Db 568 AlaThrVal-----ThrLysAlaValAlaGluLeuGlyGlyGluThrGluLysVal 584
 Qy 1103 CGAGCTCTCTCTCTAAATAATTAACGATCATGCTGCGCTATTGTTGGATTGTTGTGAGA 1162
 Db 585 ArgAspGluLeu----- 588
 Qy 1163 AACAGGCTAGATTCTCTACGGNAACTCTTAACGAACCGCAAAATTTTGGTCAAGGTA 1222
 Db 589 -----ile 589

QY 1223 AGCAGCAACTATGAGAAAGATGCTATGACGCGACCTGCGGCTTTGAGAGCTTG 1282
 DB 590 GtlyProserLeuGlyGlnGlnLeuArgAspAlaLeuIleAlaLeuGlyLeuAlaLeu 609
 QY 1283 GCAATTCATCTGCTCTATGATGAGTTGGCTTTGATGAGCAATATGCTTGAAGTCCGTA 1342
 DB 610 AlaAlaGlnLeuAlaTylLeuAlaValArgPheArgLeuLeuPheGlyThrAlaAlaVal 629
 QY 1343 TGCGCTTAATTCATGACCTTTGGCTACCTGTGCACTGCTTTGATAGCAATTTCTTT 1402
 DB 630 GlyAlaLeuAlaHisAspVal-----ValIleLeuValGlyValPheAla 644
 QY 1403 TTGAAGAAATTTCAATGATTTTGCAAGCATTGCTGCTTAATGATCTGTTATGGCGTAT 1462
 DB 645 TylLeuGlyLysProIleAspGlyValPheLeuAlaAlaLeuLeuThrValIleGlyTyr 664
 QY 1463 TCATTAAACAATCTTATGATCATCTTTGATGCTATTCGTGTA-----GATCGC 1510
 DB 665 SerValAsnAspSerValValLeuPheAspArgIleArgGlyLeuLeuGlyLysGlnArg 684
 QY 1511 CAAGGAACTGTTTACCCCTATGCGATGTTTATGATGATGCGCTTCAAAAGAGCTTC 1570
 DB 685 LysAla-----ProPheAspArgLeuThrAsnAspAlaIleLeuGlnThrLeu 700
 QY 1571 AGCCGCAAGTATGACAAAGCTCAACTCATGATGTTTGTATATGCTTTTGTATATA 1630
 DB 701 ProArgThrValAsnThrGlyMetGlyAlaValLeuIleLeuAlaSerLeuAlaIleLeu 720
 QY 1631 GCGGCTCTGCTGCTTAAATTTGCACTTTATGACCAAGGATCTTCTAGAACT 1690
 DB 721 AlaAspAspSerLeuThrAspPheAlaLeuAlaLeuIleGlyValGlyValGlyThr 740
 QY 1691 TTATGCTCTTTATATGACCACTCTGTTGTTTATGATGCTCGTAAGAAATGCGC 1750
 DB 741 TyrSerSerValPheThrAlaSerProLeuAlaIleGlyLeuHisAsnArgAspThrGly 760
 QY 1751 TCMAAA 1756
 DB 761 SerArg 762
 RESULT 13
 A11627
 protein-export membrane protein SecDF homolog l1n1562 (imported) - *Listeria innocua* (str
 C;Species: *Listeria innocua*
 C;Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001
 C;Accession: A11627
 R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker
 ; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Feihl, H.
 D.; Jones, L.M.; Kastr, U.
 Science 294, 849-852, 2001
 A;Authors: Kretz, J.; Kuhn, M.; Kunz, F.; Kurapkac, G.; Madueno, E.; Maitournam, A.; Ma
 ok, C.; Schlueker, T.; Simoes, N.; Tiller, A.; Vazquez-Boland, J.A.; Voss, H.; Wehlend,
 A.; Title: Comparative genomics of *Listeria* species.
 A;Reference number: AB1077; MUID:21537279; PMID:11679669
 A;Accession: A11627
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-754 <GB>
 A;Cross-references: GB:AL592022; PIDN:CAC96793.1; PID:G16414049; GSPDB:GN00178
 A;Experimental source: strain G1p11262
 C;Genetics:
 A;Gene: l1n1562

Alignment Scores:

Pred. No.: 4.05e-46 Length: 754
 Score: 620.50 Matches: 162
 Percent Similarity: 47.13% Conservative: 117
 Best Local Similarity: 27.36% Mismatches: 22
 Query Match: 18.73% Indels: 91
 DB: 2 Gaps: 13

US-09-868-987-1 (1-1864) x A11627 (1-754)

QY 98 TATATGTCAGACCCCTATTTTAAAGCTCCCAATGAAAAATCATGCGACGTGCTCAGG 157
 DB 196 TyrLeuSerAlaProAsnValSerSerValLeuAspThrAspLysValGluIleSerGly 215
 QY 158 AATTTTACCAAGCGTGAAGTGAACAACTGCGCTCAGATTTTAAATTTGAGGAGCTCT 217
 DB 216 SerPheThrThrGlnGlnAlaLysAspLeuAlaGluLeuLeuAsnSerGlyAlaLeuPro 235
 QY 218 TTTGTTCCAGAGTCTCAGTGAAGACGATCTCTTCATCTTGGGAAAAACAATGT 277
 DB 236 ValLysMetThrGlnValTyrSerThrSerValGlyAlaGlnPheGlyAsnAlaLeu 255
 QY 278 ACACAAGCAATTCATGACATGCTGCTGCGCAATGCTTATGTTGATGAGCGTA 337
 DB 256 GlnGlnThrIleLeuAlaGlyIleIleGlyValMetAlaIlePheIlePheMetAla 275
 QY 338 TATATGATTTGAGAGCGTCATGCTGCTGCGAGCTGTTCTTGAATCTTTGCTTATC 397
 DB 276 ValTyrArgLeuProGlyValIleAlaSerIleThrLeuValAlaTyrThrTyrLeuVal 295
 QY 398 TGGGCACTCTACAGTATTTGATGCGCACTACCTTTCAGAGACTCGCTGGATGTT 457
 DB 296 LeuLeuIleLeuSerLeuLeuAsnAlaThrLeuThrLeuProGlyIleAlaGlyLeuIle 315
 QY 458 CTGCTATGGGAGTGGCGGTAGATGCAATGTTCTTGTATCGAAAGATCCGAGAGAA 517
 DB 316 LeuGlyIleGlyMetAlaValAspAlaAsnValIleThrTyrGlnAlaGlyLysGln 335
 QY 518 TTTTATGCTCTCAAGCTCTTAAAAATCTGTAGAAAAAGATATACCAAGCTTTTGA 577
 DB 336 IleLysValGlyArgSerThrLysAlaIlePheGlnValGlyGlyLysGlnAlaPheArg 355
 QY 578 GCCATTTTGAATTCATCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 637
 DB 356 AlaIleLeuAspGlyAsnLeuThrThrLeuIleValAlaValLeuPheTyrPheGly 375
 QY 638 ACAGGCGCTTAAAGGTTGCTTGAATGATGATGATGATGATGATGATGATGATGATGAT 697
 DB 376 ThrSerSerIleLysGlyPheAlaThrValLeuIleSerIleLeuValSerPheLeu 395
 QY 698 ACGGCTCTTTCAGACTTAAATTTTCTTCATGCTG-----TGATGAAT 742
 DB 396 ThrAlaValThrGlySerArgPheLeuLeuGlyLeuLeuValLysSerAsnThrLeuAsn 415
 QY 743 AAG-----ACCAACATACACAGTTCATATGATGATGATGATGATGATGATGATGAT 784
 DB 416 AsnLysProGlyPhePheAlaValLysArgLysAspIleHisAsnLeuHisGlnGlyIle 435
 QY 785 GCGATA-----AAGCATATTTCTTGAGAGATGCAAAAA 820
 DB 436 AsnSerPheSerLeuLysThrHisPheAspArgTyrAspPheValLysHisAsnArgLeu 455
 QY 821 CTTTGGGCTGTTCTGAGAGTGTCTTTTCTTTAGGTGCTGCTCGGATTTGGAGCC 880
 DB 456 PheLeuSerIlePheAlaAlaIleValIleValGlyIleValIleLeuSerIlePheLys 475
 QY 881 TGAATTCCTGTTGGGAATGATTTTAAAGAGGATGATGATGATGATGATGATGATGATGAT 940
 DB 476 LeuAsn-----LeuGlyIleAspPheAlaSerGly-----ThrArg 487
 QY 941 GAGCATGGCATCAGCGATGTTGCTCAAAATGCGGCAAGTGTGATGATGATGATGATGATGAT 1000
 DB 488 AlaGlnValThrAlaAsnGlnThrLeuThrGlnThrGlnIleLysValAspLeuAspAla 507
 QY 1001 GCTGCTCTTTCTTGAAGACTTCGATTAACAATTTGATGATGATGATGATGATGATGATGAT 1048
 DB 508 IleAspMetProSerAsnAspIleValPheGlnGlySerGlySerLysThrAlaValAl 527
 QY 1049 -----GAAAGATCAAAATCATTTTATGATGAT 1075
 DB 528 SerTyrLysGlyThrLeuSerGlnAsnAspValAlaLysPheLysAspTyrPheGlnLys 547
 QY 1076 AAAGCTTAAGCTATACAGCAAGATGAGAGCTCTCTCTAAATTAACAGATCATGAGC 1135

```
Db 548 Lys-----TyrLysHisGluProSerIleSerThrValSerProThrVal----- 562
Qy 1136 TGGCGTTATTGTGGGATTGTTGTGAGAACAGCGCTAGATTCTCTACGGAACCTCTAAA 1195
Db 563 -----GlyLysGluLeu 566
Qy 1196 CGAAACGGAAATTTTGGTCMAAGGTAAAGCAGCAAACTATCGAAGAAATCGGTTATCAG 1255
Db 567 AlalysAsnGlyPheTrpAla----- 573
Qy 1256 GCGACCATCGGGCTTTAGGA---GCTTTGGCAATCATCTGTCTCTATGTGAGTTTGGC 1312
Db 574 -----LeuGlyValAlaSerValLeuIleValLeuIleValLeuIleValAlaValArg 588
Qy 1313 TTTGAATGGCAATATGCTTTTCAGTCGCGTATGCTTTTAAATTCATGACCTTTTGGCTACC 1372
Db 589 PheGluPheTyrMetGlyIleAlaAlaIleLeuSerLeuLeuPheAsp----- 604
Qy 1373 TGTCCAGTCTGTTTATAGCACATTTCTTTTGAAGAAATTCAAATAGATTTCGAAGCC 1432
Db 605 ---AlaPheIleIlePheIlePheSerValThrArgLeuGluValAspLeuThrPhe 623
Qy 1433 ATTGGTCTTTAAGTCTTATGCGGTATTCATTAAACAATATGATCATTTTTCAT 1492
Db 624 IleAlaValLeuThrValIleGlyTyrSerIleAsnAspThrIleValThrAlaAsp 643
Qy 1493 CGTATTCGTGAA---GATCGCAAGCAAGCTGTTTACCCTCATGCTGTTTAA----- 1543
Db 644 ArgIleArgAspIleSerMetLysMetGlnArgPheLysThrLysGluGluIleAlaAsp 663
Qy 1544 ---GTTAATGATGCCCTTCAAAGACGTTACGCGCAGCGGTAATGACAACAGCTACAACT 1600
Db 664 AlaValAsnLysAlaLeuArgGlnThrPheThrArgSerIleAsnThrIleLeuThrVal 683
Qy 1601 CTATCAGTTTGTGTTAATGCTTTTGTATAGCGCGCTCTCTGCTTTTAAATTTGCAATTT 1660
Db 684 IlePheThrValLeuAlaLeuValLeuPheGlySerGluSerIleLeuAsnPheSerIle 703
Qy 1661 ATTATGACCATAGGATCTTCTAGGAACCTTATGCTCTCTTTATATGACCACTCTG 1720
Db 704 AlaLeuLeuValGlyLeuValSerSerValPheSerSerIlePheMetAlaMetGlnLeu 723
Qy 1721 TTGTTGTTTATGTCGCGTAAGAAATCGCTCAAAA 1756
Db 724 TriPyrValPheLysAlaArgGlnLeuArgLysLys 735

RESULT 14
AG1265
protein-export membrane protein SecDF homolog lmo1527 [imported] - Listeria monocytogene
C;Species: Listeria monocytogenes
C;Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001
C;Accession: AG1265
R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Duseurget, O.; Entian, K.D.; Feihl, H.
D.; Jones, L.M.; Karst, U.
A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma
ok, C.; Schlueder, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
A;Title: Comparative Genomics of Listeria species.
A;Reference number: AB1077; MUID:21537279; PMID:11679669
A;Accession: AG1265
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-754 <GLA>
A;Cross-references: GB:NC_003210; PIDN:CAC99605.1; PID:g16410956; GSPDB:GN00177
A;Experimental source: strain EGD-e
C;Genetics:
A;Gene: lmo1527

Alignment Scores: 6.07e-46 Length: 754
Pred. No.: 618.50 Matches: 166
Score: 166
```

```
Percent Similarity: 46.79% Conservative: 111
Best Local Similarity: 28.04% Mismatches: 224
Query Match: 18.67% Indels: 91
DB: 2 Gaps: 13

US-09-868-987-1 (1-1864) x AG1265 (1-754)
Qy 98 TATATGTCAGCAGCCCTATTTTAAACGTCCCATTTGAAATATCATGCCAGTGTCTCAGGG 157
Db 196 TyrLeuSerAlaProAsnValSerSerValLeuAspThrAspLysValGluIleSerGly 215
Qy 158 AATTTTACCACCGTCGAGTCAGCAAACTCGCCTCAGATTTAAATCTCGAGCGATGCT 217
Db 216 SerPheThrThrGluGluAlaLysAspLeuAlaGluLeuLeuAsnSerGlyAlaLeuPro 235
Qy 218 TTTGTTCCCGAGGTTCTCAGTGAAGAGACGATCTCTTCATCTTGGGAAAAACAATGT 277
Db 236 ValLysMetLysGluValTyrSerThrSerValGlyAlaGlnPheGlyGlnAspAlaLeu 255
Qy 278 ACACAGGCATTTATCTCAGCATCTGTGGCTTGGCAATGCTTATTTGTTGATCAGCGTA 337
Db 256 GlnGluThrIleLeuAlaGlyIleIleGlyValIleAlaIlePheIlePheMetMetAla 275
Qy 338 TATATAGATTTGAGCGCTACGCTTCGGAGCTGTTCTTCTGAATCTTTTCTTATC 397
Db 276 ValTyrArgLeuProGlyValIleAlaSerIleThrLeuValAlaTyrThrTyrLeuVal 295
Qy 398 TGGCAGCTCTACAGTATTGAGTGGCCACTCACCTTGTGAGGACTCGCTGGGATTTGT 457
Db 296 LeuLeuLeuLeuSerLeuLeuAsnAlaThrLeuThrLeuProGlyIleAlaGlyLeuIle 315
Qy 458 CTTGCTATGCGGATGCGCTAGATGCAAAATGTTCTTGTATTGAAAGAAATCCGAGAGAA 517
Db 316 LeuGlyIleGlyMetAlaValAspAlaAsnValIleThrTyrGluArgIleLysGluGlu 335
Qy 518 TTTTATTTGCTCAAGTCTTAAAAATCTGTAGAAAAAGGATATACCAAGCTTTTGA 577
Db 336 IleLysValGlyArgSerThrLysAlaAlaPheGluValGlyGlyLysGluAlaPheArg 355
Qy 578 GCATTTTGTATTTTCTTACCTGACTACAGTATTGCGCTCAGCACTCTTTCTTCTTCTAGAT 637
Db 356 AlaIleLeuAspGlyAsnLeuThrThrLeuIleValAlaAlaValLeuPheTyrPheGly 375
Qy 638 ACAGGCGCTATTAAAGGGTTTGTCTTGACATTTATTAGGAATTTTCTCTCAATGTTT 697
Db 376 ThrSerSerIleLysGlyPheAlaThrValLeuIleIleSerIleLeuValSerPheLeu 395
Qy 698 ACGGCTCTTTTCATGACTAAATTTTCTTCATGCTG-----TGGATGAAT 742
Db 396 ThrAlaValTrpGlySerArgPheLeuLeuGlyLeuLeuValLysSerAsnTrpLeuAsn 415
Qy 743 -----AAGACCCCAACATACACAGTTGTCATATGATG 772
Db 416 AsnLysProGlyPhePheAlaValLysArgLysAspIleHisAsnLeuHisGluGlyIle 435
Qy 773 AATAAGTTCGTGGGATAAAGCAT-----GATTTCTTGAGAGGATGCAAAAAA 820
Db 436 AsnSerPheSerLeuLysThrHisPheAspArgPheAspPheValLysHisHisArgLeu 455
Qy 821 CTTTGGGCTGTTTCTCGAAGTGTGTTTCTTTTAGTTTGGCTTCTCTCGGGTTTGGAGCC 880
Db 456 PheLeuSerIlePheAlaIleValIleValIleValIleLeuSerIlePheArg 475
Qy 881 TGGAAATTCGTTTGGGAATGGAATTTTAAAGGAGGATGATGCCTTTTACCTTTTATCCAAA 940
Db 476 LeuAsn-----LeuGlyIleAspPheAlaSerGly-----ThrArg 487
Qy 941 GAGCATGGCATCAGCGATGTTGCTCAAAATGCGTGGCAAGCTTGTGCATAAACTACAGAA 1000
Db 488 AlaGluValThrAlaAsnGlnThrLeuThrGluThrGlnIleLysLysAspLeuAspThr 507
Qy 1001 GCTGGTCTTTCTTCTAGAGACTTTCGGTATTCAAAACATTTTGGATCTTCA----- 1048
Db 1001 GCTGGTCTTTCTTCTAGAGACTTTCGGTATTCAAAACATTTTGGATCTTCA----- 1048
```


Db 508 ILeaspMetProSerAspAPILeValPheGlnGlySerGlySerThrAlaVal 527
 QY 1049 -----GAAAGATCAAAATCTATTATTTAGCAT 1075
 Db 528 SerTyrIysGlyThrLeuSerGlnAsnAspValAlaIysPheIysAsnTyrPheGluAsp 547
 QY 1076 AAACCTTTAGCTPACACAGATACAGACCTCTCTCCCAAAATTAATACATCATAGC 1135
 Db 548 Lys-----TyrLysHsGlnProSerIleSerThrValSerProthVal----- 562
 QY 1136 TGGCGTTATGTGGAGATTGTGTCAAAACAGCGCTAGATTCTCTACGAAACTTAA 1195
 Db 563 -----GlyLysGluLeu 566
 QY 1196 CGAAACGCAAAATTTGGTCAAAAGTAAGCAGCAAACTATCGAAGAAATGCCATTATAG 1255
 Db 567 AlalysAsnGlyPheTrpAla----- 573
 QY 1256 GCGACCATCGGGCTTTTGA---GCTTGGCAATCATCTGCTATGATGAGTTTGGC 1312
 Db 574 -----LeuGlyValAlaSerValLeuIleValLeuTyrIleAlaValArg 588
 QY 1313 TTGGAATGCAATATAGCTTTAGTCCGTATGCGCTTATATCATGACCTTTGGCTAC 1372
 Db 589 PheGluPheTyrMetGlyIleAlaAlaIleLeuSerLeuLeuPheAsp----- 604
 QY 1373 TGTGAGTCTGTGTTATAGACATTTCTTTTGAAGAAATTCAAATAGATTGGAC 1432
 Db 605 ---AlaPheIleIlePheIlePhePheSerValThrArgLeuGluValAspLeuThrPhe 623
 QY 1433 ATGGTGTCTTAATAGCTGTTATGGGATTCATTAACATTAATCTTATCATCTTTTAT 1492
 Db 624 IleAlaIleValLeuThrValIleGlyTyrSerIleAsnThrIleValThrAlaAsp 643
 QY 1493 CGTATTCGTGA---GATCGCCACAGCAACCTGTACCCCTATGATGTTTAA----- 1543
 Db 644 ArgIleAspAPILeSerMetLysMetGlnArgPheLysThrLysGluGlnIleAlaAsp 663
 QY 1544 ---GTTAATGATGCCCTTCAAAAGACGTTACAGCCGAGGTATACACACAGCTCACT 1600
 Db 664 AlaValAsnLysAlaLeuArgGlnThrPheThrArgSerIleAsnThrIleLeuThrVal 683
 QY 1601 CTATCAGTTTGTCTTATAGCTTTTGTATAGCGCGCTCCTGCTTAAATTTGATTT 1660
 Db 684 IlePheThrValLeuAlaLeuValLeuPheGlySerGluSerIleLeuAsnPheSerIle 703
 QY 1661 ATTATGACCATAGGATTTCTTCTAGAACTTTATGCTCTCTTATATGACACACCTCTG 1720
 Db 704 AlaLeuLeuValGlyLeuValSerSerValPheSerSerIlePheMetAlaMetGlnLeu 723
 QY 1721 TTGTTGTTATGTCCTTAAAGAAATGCTCAAA 1756
 Db 724 TrpTyrValPheLysAlaArgGlnLeuArgLysLys 735
 RESULT 15
 C90594
 protein-export membrane protein secd [imported] - Mycoplasma pulmonis (strain UAB CTIP)
 C/Species: Mycoplasma pulmonis
 C/Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 03-Aug-2001
 C/Accession: C90594
 R/Chambud, I.; Hellig, R.; Ferris, S.; Barde, V.; Samson, D.; Galisson, F.; Moszer, I.;
 Nucleic Acids Res. 29, 2145-2153, 2001
 A/Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma pul
 A/Reference number: A9512; MUID:21267165; PMID:11353084
 A/Accession: C90594
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-873 <KUR>
 A/Cross-references: GB:AL445566; PID:gl4090074; PIDN:CAC13632.1; GSPDB:GN00153
 A/Experimental source: strain UAB CTIP
 C/Genetics:
 A/Gene: MYPU 6590
 A/Genetic code: SGC3

Alignment Scores:
 Pred. No.: 1,688-37 Length: 873
 Score: 522.50 Matches: 143
 Percent Similarity: 47.778 Conservative: 135
 Best Local Similarity: 24.578 Mismatch: 195
 Query Match: 15.784 Indels: 109
 DB: 2 Gaps: 18
 US-09-868-987-1 (1-1864) x C90594 (1-873)
 QY 149 GTCTCAGGAAATTTATCCACCGTGAAGTGAACAACTCGCTCAGATTAAATCTGCA 208
 Db 298 IleThrGlyAsnLeuThrAlaSerSerAlaLysGlnLeuAlaLeuAspValAsnPheGly 317
 QY 209 GCGATGCTTTGTGTCCTCCCGAGTTCACAGGAAGAGACATCTCTCTGATCTTGGGAA 268
 Db 318 SerGlyAspTyrThrLeuGlnLeuLeuSerSerArgPheValAspAlaThrLeuGlyThr 337
 QY 269 AAACATGTACACAGGACATTTATCTCAGCATGCTGTGGCAATGCTTATGTTTGG 328
 Db 338 GlnSerPheAsnTyrAlaIleIleAlaGlyIle-----IleSerPheIleValIleAla 355
 QY 329 ATGACGTATATATATAGATTGGA-----GGCGTATCGCTTGGAGCTGTTCTTCTG 382
 Db 356 LeuIleMetIleTyrAsnTyrGlyValLeuGlyIleAlaSerSerLeuSerIleGlyLeu 375
 QY 383 AATCTTTGCTTATCTGCGACGCTTACAGTATTTGGATGCGCACCTCACCTTGTCA 442
 Db 376 TyrIleLeuLeuSerIleThrValPheThrTrpLeuAsnGlyLeuTyrSerProAlaThr 395
 QY 443 CTCGCTGGATTTGTTCTTGTATGAGGAGCGGTATGATCAAAATGTTCTGTATTCGA 502
 Db 396 IleAlaIleValIleAlaIleGlyMetSerValAspAlaAsnIleIleThrPheGlu 415
 QY 503 AGAATCGGAGAGAAATTTATGTCTCAAAATCTTAAATAATGTGAAGAAAGATAT 562
 Db 416 ArgIleLysGlnGlnLeuLysTyrGlyAspSerLeuLysAlaThrLysAsnAlaHis 435
 QY 563 ACCAAGCTTTGGAGCCATTTTATGATTTTAACTTAACTGATACAGTATGGCTCAGCACT 622
 Db 436 ArgPheSerLeuSerThrIleLeuAspSerAsnIleThrThrLeuIleValSerGlyIle 455
 QY 623 CTTTCTTCTCAGATACAGGCGCTTAAAGGTTGCTTGAATTTAGATT 682
 Db 456 LeuThrTyrPheGlyThrLysLeuIleArgGlyPheSerIleSerLeuIleLeuSerIle 475
 QY 683 TTCCTTCATGTTTACGGCTTTTCAATGATTAATTT----- 721
 Db 476 IlePheIleLeuPheValMetLeuIleIleThrArgMetAlaThrSerLeuLeuIleGly 495
 QY 722 -----TTCCTCATGCTGTGATGAATTAAGCCACACATACACGTTGCATATGATAT 775
 Db 496 ThrGlyPhePhe-----AspAsnArgLeuTyrLeu----- 505
 QY 776 AAGTTCGGGAGATTAAGCATGATTTCTTGAGAGATGCAAAAATCT----- 823
 Db 506 -----IleGlyValGlyArgSerPhePheThrLysGlyLysIleIleThrPhePhe 523
 QY 824 -----TGGCGTGT---TCGGAAGTGTTTTCTT 850
 Db 524 GluLysLeuAsnTyrPheLysLeuSerTyrPheIleLeuPhePheGlyValPheVal 543
 QY 851 TTAGGTCGCTGCTCTCGGGTTTGAGCCCTGGAATTCCTTTTGGCA----- 898
 Db 544 LeuIleAlaIleIleIle---TyrSerIlePheIleGlyLeuAsnGlyAsnPheLeuSer 562
 QY 899 -----ATGATTTTAAAGAGGAGTATGCTTATCTTAAATCCAAA--- 940
 Db 563 GlyPheGluArgSerLeuGlnPheGlnGlyThrAsnIleSerPheGluLysSerSer 582
 QY 941 -----GAGCATGCAATCAGGATGCTGCTCAAAATGCGTGGCAAAATTTGTCATTAAC 994

Db 583 IleAspSerLeuSerLeuAlaGluAlaGlnGluLeuLeuLeuValAsnLeuLeu 602
Qy 995 CAGGAAGCTGCTCTTCTCTAGAGACTTCGCTATCAACATTTGGATCTTCAGAAAG 1054
Db 603 AsnThrAlaAsn-----GlySerGluAsnLeu 611
Qy 1055 ATCAAAATCTATTTAGTGATAAGCTTTAGCTATCTAGCAGATACAGCCCTCTCTC 1114
Db 612 IleThrSerGlyAlaGlnAspPheIleAsnAlaLysAsnGlnAlaArgSerAsn--- 630
Qy 1115 CTAATAATTAACGATCATGAGCTGGCTTATGTGGATTGTGTGAGAAACAGCCCTAGA 1174
Db 631 -----TyrValValSerLeuGlnThrLys 638
Qy 1175 TTTCTCTACGAAACTCTTAACGAAAC-----GCAAAATTTTGTGCA 1216
Db 639 GlnValPheAspGlnAspGlnIleAsnTyrIleIleGlnGluValSerLysThrPheSer 658
Qy 1217 AAGTAAAG-----ACCAACTATCAAGAA 1243
Db 659 AsnIleSerSerThrProPheIleValSerThrAsnGluAlaSerSerLeuValLysAsn 678
Qy 1244 ATGGCTTATCAGGCGACCATCGGCTTTAGGAGCTTTGGCAATCATCTCTCTATGTG 1303
Db 679 SerIleValSerIleSerIleGlyIleValGly-----IleValLeuTyrThr 694
Qy 1304 AGTTGGCTTTGAATGCAATATGCTTTAGTCGCCGTATGCGCTTTAATTCATGACCTT 1363
Db 695 IlePheArgLeuLysTyrThrPheSerIleAlaIleValAlaLeuLeuHisAspIle 714
Qy 1364 TTGGCTACCTGTGAGCTGTTTATAGCACATTTCTTTTGAAGAAATTCAAATAGAT 1423
Db 715 -----ValPheValIleAlaPheIleIlePheArgLeuGluLeuSer 729
Qy 1424 TTGAAGCCATTGGTCTTAACTGCTATTTGGGTATTTCAATAAACAATCTTTGATC 1483
Db 730 ProIlePheValAlaIleLeuSerIleIleGlyTyrSerIleAsnAspThrIleVal 749
Qy 1484 ATTTTGTATGCTATTCGTGAAGATCGCAAGCAACCTGTTT-----ACCCCTATG 1534
Db 750 ThrPheAspArgIleLysGluIleMethHisSerGluTyrTyrHisGlnLysIleAspGln 769
Qy 1535 CATGTTTAGTT-----AATGATGCCCTTCAAGAGCTTCAGCGCACCGTAAATG 1585
Db 770 LysValLeuIleAsnIleThrAsnLysAlaIleLysAspThrIleLysArgSerMetLeu 789
Qy 1586 ACAACAGCTCAACTCTATCATGTTTGTGTTATGCTTTGTTTATAGGCGCTCTCTGTC 1645
Db 790 ThrThrIleThrThrLeuValThrIleValValLeuLeuSerPheArgLysThrGlu 809
Qy 1646 TTTAATTTTGCAATTATATGACCATAGGATTTCTCTAGGAATTTATCTCTCTCTTAT 1705
Db 810 ValAsnPheAsnIleAlaLeuPheValGlyLeuIleSerGlyThrTyrSerSerIlePhe 829
Qy 1706 ATTGCA 1711
Db 830 IleAla 831
RESULT 16
S76266
hypothetical protein - Synechocystis sp. (strain PCC 6803)
C:Species: Synechocystis sp.
A:Variety: PCC 6803
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 16-Jun-2000
C:Accession: S76266
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; O. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yasuda DNA Res. 3, 109-136, 1996
C:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis s.
A:Reference number: S74322; MUID:97061201; PMID:8905231
A:Accession: S76266
A>Status: preliminary

A:Molecule type: DNA
A:Residues: 1-472 <KAN>
A:Cross-references: EMBL:D64000; GB:AB001339; NID:g1001484; PID:g1001493
C:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C:Superfamily: protein export membrane protein secD
Alignment Scores:
Pred. No.: 1.43e-30 Length: 472
Score: 443.00 Matches: 89
Percent Similarity: 62.39% Conservative: 57
Best Local Similarity: 38.03% Mismatches: 82
Query Match: 13.38% Indels: 6
DB: Gaps: 1

US-09-868-987-1 (1-1864) x S76266 (1-472)

Qy 50 CAATATTTCTCAAAACCGTGATGCGGTATGCTGTAGTATTGACGGTTATATGTCAGC 109
Db 213 GlnAlaValAlaGlyThrGlyArgSerLeuGlyValPheLeuAspAsnAspLeuIleSer 232
Qy 110 AGCCCTATTTTAAACGCTCCATTTGAAAAAT-----CATGCCAGTGTC 151
Db 233 AlaProValValGlyValGluPheAlaAsnThrGlyIleThrGlyAlaAlaValIle 252
Qy 152 TCAGGGAATTTTACCACCGTGAAGTGCAGCAAACTGCCCTCAGATTTAAATCTGGAGCG 211
Db 253 ThrGlyAsnPheThrIleAspThrAlaAsnAspLeuAlaValGlnLeuArgGlyGlySer 272
Qy 212 ATGTCCTTTGTTCCCGAGGTTCTCAGTGAAGAGACGATCTCTCTGATCTTGGGAAAAA 271
Db 273 LeuProPheProValGluValValGluAsnArgThrValGlyAlaThrLeuGlyGlnGlu 292
Qy 272 CAATGTACACAAGCGCATTTCTCAGCATGCTGTGGCTTGGCAATGCTTATTTGTTTGTG 331
Db 293 SerIleArgArgSerLeuValAlaGlyPheValGlyLeuValLeuValPheMet 312
Qy 332 AGCGTATATTATAGATTGGAGCGTCATCGCTTCGGAGCTGTTCTTCTGAATCTTTTG 391
Db 313 AlaValTyrTyrArgLeuProGlyIleValAlaAspIleSerLeuMetIleTyrAlaVal 332
Qy 392 CTATCTGGGCACCTCTACAGTATTGATGTCGCCCATCTACCTTGTCCAGGACTCGCTGGG 451
Db 333 LeuThrLeuAlaAlaPheAlaLeuValGlyValThrLeuThrLeuProGlyIleAlaGly 352
Qy 452 ATGTCTTCTCTAGTGGGATGGCGTAGATGCAAAATGTTCTTGTATTGAAAGAAATCCGA 511
Db 353 PheIleLeuSerIleGlyMetAlaValAspAlaAsnValLeuIlePheGluArgThrArg 372
Qy 512 GAGGAATTTTATGCTCAAGCTTTAAATAATCTGTAGAAAAAGGATATACCAAGCT 571
Db 373 GluGluLeuArgAlaGlyAsnThrLeuTyrArgSerValGluAlaGlyPheArgAla 392
Qy 572 TTGGAGCCATTTTGTACTTACTTACCTACAGTATGTCCTCAGGCTCAGCCTCTTTCTTC 631
Db 393 PheSerSerIleLeuAspSerAsnValThrLeuIleAlaCysAlaLeuPheTrp 412
Qy 632 CTAGATACAGCGCTATTAAAGGTTTGTCTTGCATTTTGTAGGAATTTTCTCTTCA 691
Db 413 PheGlySerGlyLeuValLysGlyPheAlaLeuThrLeuAlaIleGlyValMetValSer 432
Qy 692 ATGTTACGCTCTTTTTCATGACTAAATTTTCTTCTCATGCTG 733
Db 433 LeuPheThrAlaLeuThrCysSerArgThrLeuLeuLeuVal 446

RESULT 17

A:11821
protein-export membrane protein [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp.
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002
C:Accession: A11821
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, S.

DNA Res. 8, 205-213, 2001
 A/Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium *Ana*
 A/Reference number: AB1807; MUID:21595285; PMID:11759840
 A/Accession: A11821
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-474 <KIR>
 A/Cross-references: GB:BA000019; PIDN:BA877645.1; PID:917135099; GSPDB:GN00179
 A/Experimental source: strain PCC 7120
 C/Genetics:
 A/Genetic:
 C/Superfamily: protein export membrane protein secD

Alignment Scores:
 Pred. No.: 5.89e-30 Length: 474
 Score: 436.00 Matches: 89
 Percent Similarity: 60.61% Conservative: 51
 Best Local Similarity: 38.53% Mismatches: 85
 Query Match: 13.16% Indels: 6
 DB: Gaps: 1

US-09-868-987-1 (1-1864) x A11821 (1-474)

QY 59 GCAACCGTGAAGGCGTATGCGTATGATGACGGTTATATGTCAGACCCCTAT 118
 Db 214 AAGGlyThrGlyArgSerIleGlyIlePheLeuAspAsnGluLeuIleSerAlaProAsn 233
 QY 119 TTAAACGTCCATGTAATAAT-----CATGCCAGTGTCTCAGGAAA 160
 Db 224 ValGlyIleGluPheAlaSerThrGlyIleThrGlyIleSerAlaValIleThrGlyArg 253
 QY 161 TTTACCCACCGTGAAGGCGTATGCGTATGATGACGGTTATATGTCAGACCCCTAT 220
 Db 254 PheThrAlaGlnAlaAsnAspLeuGlyValGlnLeuArgGlyValAlaLeuProVal 273
 QY 221 GTTCCGAGGCTTCAGTAAGAGACGATCTCTGATCTTGGAAAAACAATGTACA 280
 Db 274 ProValGluIleAlaGluIleArgThrValGlyAlaThrLeuGlyIleAspSerIleAsn 293
 QY 281 CAAGCATATATCTCAGATGCTGTGCGTTCGCAATGCTTATGTTGATGACCGTATAT 340
 Db 294 SerSerIleTyrAlaGlyLeuGlyIleLeuThrLeuValLeuIlePheMetValValTyr 313
 QY 341 TATGATTTGAGGCGTATGCGTATGCGTATGATGACGGTTATATGTCAGACCCCTAT 400
 Db 314 TyrArgLeuProGlyLeuIleAlaAspIleSerLeuIleIleTyrAlaIleLeuThrTrp 333
 QY 401 GCAGCTTACAGTATTTGATGCGCCACGACCTGTGACGACGCTGGGATTTGCTT 460
 Db 334 AlaSerPheAlaLeuLeuGlyIleThrLeuThrLeuProGlyIleAlaGlyIleLeu 353
 QY 461 GCTATGGGATGCGCGTATGCGCAATGTTCTTATGTAAGAATCCAGAGGAATTT 520
 Db 354 SerIleGlyMetAlaValAlaAspAlaAsnValLeuIlePheGluArgThrArgGluLeu 373
 QY 521 TTATTTGCTCAAGCTTTAAAAATCTGTAGAAAAAGATATCCAGAGCTTTTGAGACC 580
 Db 374 GlnAlaGlyIleSerLeuTyrSerValGlnSerGlyPheTyrArgAlaPheSerSer 393
 QY 581 ATTTTGTGATTTGATGCGTATGCGTATGCGTATGCGTATGCGTATGCGTATGCGTAT 640
 Db 394 IleLeuAspGlyAsnValThrValIleAlaCysAlaAlaLeuPheTrpLeuValAla 413
 QY 641 GGGCTATTAAGGCTTTGCTTGAATGATTTAGAAATTTTCTTCAATGTTATAG 700
 Db 414 GlyLeuValIleGlyPheAlaLeuThrIleAlaLeuGlyValAlaValSerMetPheSer 433
 QY 701 GCTCTTTCATGACTTAAATTTTCTTCAATGCTG 733
 Db 434 AlaValThrCysSerArgThrLeuMetPheLeu 444

RESULT 18
 F64713

protein-export membrane protein - *Helicobacter pylori* (strain 2695)
 C/Species: *Helicobacter pylori*
 C/Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jun-2000
 C/Accession: F64713
 R/Tomb, J.F.; White, O.; Kariavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.
 Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKen
 son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey,
 Nature 388, 539-547, 1997
 A/Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser,
 A/Title: The complete genome sequence of the gastric pathogen *Helicobacter pylori*.
 A/Reference number: A64520; MUID:97394467; PMID:9252185
 A/Accession: F64713
 A/Status: preliminary; nucleic acid sequence not shown; translation not shown
 A/Molecule type: DNA
 A/Residues: 1-503 <TOM>
 A/Cross-references: GB:AE00652; GB:AE00511; NID:G2314720; PIDN:AA08588.1; PID:G2314
 C/Genetics:
 A/Start codon: GTG
 C/Superfamily: protein export membrane protein secD

Alignment Scores:
 Pred. No.: 1.64e-29 Length: 503
 Score: 431.00 Matches: 99
 Percent Similarity: 60.83% Conservative: 47
 Best Local Similarity: 41.25% Mismatches: 92
 Query Match: 13.01% Indels: 2
 DB: Gaps: 2

US-09-868-987-1 (1-1864) x F64713 (1-503)

QY 47 GGACATATTTCT---GCAACCGTGAAGGCGTATGCGTATGATGACGGTTATATG 103
 Db 252 GlyAspPheSerGlyAlaAsnValGlyLysArgMetAlaIleValLeuAspAsnLysVal 271
 QY 104 GTCAGACCGCTATTTTAAACGTCCCATG---AAAATATGCTCAGTGTCTCAGGAAA 160
 Db 272 TyrSerAlaProValIleArgGluArgIleGlyGlySerGlyGlnIleSerGlyAsn 291
 QY 161 TTTACCCACCGTGAAGGCGTATGCGTATGCGTATGATGACGGTTATATG 220
 Db 292 PheSerValAlaGlnAlaSerAspLeuAlaIleAlaLeuArgSerGlyAlaMetSerAla 311
 QY 221 GTTCCGAGGCTTCAGTAAGAGACGATCTCTGATCTTGGAAAAACAATGTACA 280
 Db 312 ProIleGlnValLeuGluLysArgIleIleGlyProSerLeuGlyLysAspSerValLys 331
 QY 281 CAAGCATATATCTCAGATGCTGTGCGTTCGCAATGCTTATGTTGATGACCGTATAT 340
 Db 332 ThrSerIleIleAlaLeuValGlyGlyPheIleLeuValMetGlyPheMetValLeuTyr 351
 QY 341 TATGATTTGAGGCGTATGCGTATGCGTATGATGACGGTTATATG 400
 Db 352 TyrSerMetAlaGlyValIleAlaCysLeuAlaLeuValAlaAsnLeuPheLeuIleVal 371
 QY 401 GCAGCTTACAGTATTTGATGCGCCACGACCTGTGACGACGCTGGGATTTGCTT 460
 Db 372 AlaAlaMetAlaIlePheGlyAlaThrLeuThrLeuProGlyMetAlaGlyIleValLeu 391
 QY 461 GCTATGGGATGCGCGTATGCGCAATGTTCTTATGTAAGAATCCAGAGGAATTT 520
 Db 392 ThrValGlyIleAlaValAlaAspAlaAsnIleIleLeuAsnGluArgIleAsnGluValLeu 411
 QY 521 TTATTTGCTCAAGCTTTAAAAATCTGTAGAAAAAGATATCCAGAGCTTTTGAGACC 580
 Db 412 ArgGluAsnGluGlyIleAlaValAlaIleIleIleValGlyIleAsnAlaSerArgAla 431
 QY 581 ATTTTGTGATTTGATGCGTATGCGTATGCGTATGCGTATGCGTATGCGTATGCGTAT 640
 Db 432 IlePheAspSerAsnIleThrSerLeuIleAlaSerValLeuLeuTyrAlaTyrGlyThr 451
 QY 641 GGGCTATTAAGGCTTTGCTTGAATGATTTAGAAATTTTCTTCAATGTTATAG 700
 Db 452 GlyAlaIleGlySerPheAlaLeuThrThrGlyIleGlyIleLeuAlaSerIleIleThr 471

QY 218 TTTGTTCCGAGGTTCTCAGTGAAGAGAGGATCTCTTCTGATCTTGGGAAAAACAATGT 277
Db 437 AlaProMetTyrPheAlaGluGluArgThrIleGlyProSerLeuGlyAlaAspAsnIle 456
QY 278 ACACAAGCATTATCTACAGCATGCTGTGGCTTGGCAAGCTTATTGTTTTCATGAGCGTA 337
Db 457 AlalysGlyIleAspAlaSerLeuTrpGlyMetLeuPheValSerLeuPheIleIleVal 476
QY 338 TATTATAGATTGGAGCGTCATCGCTCGGAGAGCTGTCTTCTGAAATCTTTTCTTATC 397
Db 477 IleTyrArgPhePheGlyValIleAlaThrValAlaLeuAlaPheAsnMetValMetLeu 496
QY 398 TGGCAGCTCTACAGTATTGGATGGCGCCATCCTGCTGAGCACTGCTGCGGATGTT 457
Db 497 ValAlaLeuMetSerIleLeuGlyAlaThrLeuThrLeuProGlyIleAlaGlyIleVal 516
QY 458 CTGTGATGGGGATGGCGCTAGATGCAATGTTCTGTATTGAAAGAAATCCGAGAGGAA 517
Db 517 LeuThrMetGlyMetAlaValAspAlaAsnValIleIlePheSerArgIleArgGluGlu 536
QY 518 TTTTATTGTTCTAAAGTCTTAAAGTCTTAAAGTCTTAAAGTCTTAAAGTCTTAAAGT 577
Db 537 LeuAlaAsnGlyMetSerValGlnArgAlaIleIleGluGlyPheAsnArgAlaPheThr 556
QY 578 GCCATTTTGTATTCTAAGGTTGCTTGAATGTTGTTGTTGTTGTTGTTGTTGTTGTTG 637
Db 557 AlaIleLeuAspAlaAsnLeuThrSerLeuValGlyIleLeuThrAlaMetGly 576
QY 638 ACAGGGCTATTAAAGGTTGCTTGAATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 697
Db 577 ThrGlyProValIleGlyPheAlaValThrMetSerLeuGlyIleIleThrSerMetPhe 596
QY 698 ACGGCTCTTTTCATGACTAAATTTTCTTCATGCTGTGGTGAATTAAGCCCAACATACA 757
Db 597 ThrAlaIleMetValThrArg-----Lys 616
QY 758 CAGTGTGATATGATGATTAAGTCTGTGGGATTAAGCATGATTTCTTGAGAGGATGCAA 817
Db 604 -----AlaMetValAsnLeuIlePheGly---GlyArgAspPhe-----Lys 616
QY 818 AAATTTGG 826
Db 617 LysLeuTrp 619
RESULT 23
I64056
secretion protein secD - Haemophilus influenzae (strain Rd KW20)
C:Species: Haemophilus influenzae
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 28-Jul-2000
C:Accession: I64056
R;Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.
; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J.
; D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.
Science 269, 496-512, 1995
A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter, A.
A:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
A:Reference number: A64000; MUID:95350630; PMID:7542800
A:Accession: I64056
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-616 <TIGR>
A:Cross-references: GB:U32710; GB:L42023; NID:g1573200; PIDN:AAC21908.1; PID:g1573205; T
C:Genetics:
A:Gene: secD
C:Superfamily: protein export membrane protein secD
F:Keywords: inner membrane; protein export; transmembrane protein
F:10-30/Domain: transmembrane #status predicted <TM1>
F:457-473/Domain: transmembrane #status predicted <TM2>
F:478-498/Domain: transmembrane #status predicted <TM3>
F:503-519/Domain: transmembrane #status predicted <TM4>
F:565-581/Domain: transmembrane #status predicted <TM5>
F:587-606/Domain: transmembrane #status predicted <TM6>

Alignment Scores: 1.61e-27 Length: 616
Pred. No.: 408.50 Matches: 87
Percent Similarity: 65.85% Conservative: 48
Best Local Similarity: 42.44% Mismatches: 59
Query Match: 12.33% Indels: 11
DB: Gaps: 2
US-09-868-987-1 (1-1864) x I64056 (1-616)
QY 137 AATCATCCAGTGTCTCAGCGGAAATTT-----ACC 166
Db 395 AsnValAlaThrIleGlnGlyArgPheGlySerAsnPheGlnIleThrGlyValAspSer 414
QY 167 CACCGTGAAGTGAACAACTCGCTCAGATTTAAATCTCGAGCGATGTCTTTTGTTCCTCC 226
Db 415 IleAlaGluAlaHisAsnLeuSerThrLeuLeuLysSerGlyAlaLeuIleAlaProIle 434
QY 227 GAGTTCTCAGTGAAGAGAGATCTCTTCTGATCTTGGGAAAAACAATGTACACAGGC 286
Db 435 GlnIleValGluGluArgThrIleGlyProSerLeuGlyAlaGlnAsnValGluGlnGly 454
QY 287 ATTATCTCAGCATGCTGTGCTGGCAATGCTTATTTGATGAGCGTATATTATAGA 346
Db 455 IleAsnAlaSerLeuTrpGlyLeuValAlaValIleAlaPheMetLeuPheTyrTyrLys 474
QY 347 TTTGGAGCGTCATCGCTCGGAGAGCTGTCTTCTGAATCTTTTGTCTTATCTGGCGAGCT 406
Db 475 MetPheGlyValIleAlaSerPheAlaLeuValIleAsnIleValLeuLeuValGlyLeu 494
QY 407 CTACAGTATTG---GATGCGCACCTCCTTGTGAGACTCGCTGGGATTTCTTGTCT 463
Db 495 MetSerIleLeuProGlyAlaThrLeuSerMetProGlyIleAlaGlyIleValLeuThr 514
QY 464 ATGGGGATGCGCTAGTGAATGTTCTTGTATTCGAAAGATCCGAGAGGATTTTA 523
Db 515 LeuGlyMetSerValAspAlaAsnValIleIlePheGluArgIleLysGluIleArg 534
QY 524 TTGTCTCAAAAGTCTTAAAGATCTGTAGAAAAGATATACCAAGGCTTTTGGAGCCATT 583
Db 535 AsnGlyArgSerIleGlnAlaIleAsnGluGlyTyrAsnGlyAlaPheThrSerIle 554
QY 584 TTGTATCTAATCTGACTACAGTATTGGCTCAGCATCTTTTCTTCTTCTAGATACAGGG 643
Db 555 PheAspAlaAsnLeuThrThrIleLeuThrAlaIleIleLeuTyrAlaValGlyThrGly 574
QY 644 CCTATTAAAGGTTGCTTGTGACATTTAGGATTTCTTCAATGTTTACCGCT 703
Db 575 ProlleGlnGlyPheAlaIleThrLeuSerLeuGlyValAlaIleSerMetPheThrAla 594
QY 704 CTTTTCATGACTAAA 718
Db 595 IleThrGlyThrArg 599
RESULT 24
H81312
protein-export membrane protein Cj1093c [imported] - Campylobacter jejuni (strain NCTC 11
C:Species: Campylobacter jejuni
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 03-Jun-2002
C:Accession: H81312
R;Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chilling
C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVliet, A.; Whitehead, S.; Barrell
Nature 403, 665-668, 2000
A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hyp
A:Reference number: AB1250; MUID:20150912; PMID:10688204
A:Accession: H81312
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-526 <PAR>
A:Cross-references: GB:AL139077; GB:AL111168; NID:g6968444; PIDN:CB873348.1; PID:g696852;
A:Experimental source: serotype O2, strain NCTC 11168
C:Genetics:
A:Gene: secD; Cj1093c

protein-export membrane protein SecD [imported] - Caulobacter crescentus
B87496

CISpecies: Caulobacter crescentus
CDate: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
CAccession: B87496
R.Nieterman, W.C.; Feldblum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolor
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A.Title: Complete Genome Sequence of Caulobacter crescentus.
A.Reference number: AB7249; MUID:21173698; PMID:11259647
A.Accession: B87496
A.Status: preliminary
A.Molecule type: DNA
A.Residues: 1-532 <STO>
A.Cross-references: GB:AB005673; NID:g13423458; PIDN:AAK23966.1; GSPDB:GN00148
C.Geneatics:
A.Gene: CC1991

Alignment Scores:

Pred. No.:	5,39e-26	Length:	532
Score:	391.00	Matches:	85
Percent Similarity:	60.34%	Conservative:	58
Best Local Similarity:	35.86%	Mismatch:	90
Query Match:	11.81%	Indels:	4
DB:	2	Gaps:	3

US-09-868-987-1 (1-1864) x B87496 (1-532)

DQ 62 AACCGTAGGCGCTAGTGGCTGTAGTGATTGAACGGTTATATGTCACGCCCTTAATTTTA 121
|||||
|||
|
Db 295 AsnValIGlyAsrPhelAlaIleValLeuAspGIyArgIleIleSerAlaProThrIle 314
|||||
|||
|
DQ 122 AAC---GTCCCAATTGAAAATCATGCACAGTCAGGAAATTTAACCCACCGTGAAAG 178
|||||
|||
|
Db 315 AsnGIyAlaIleLeuLylglyserGIyllleImGIySerPheThrlaIgluSerAla 334
|||||
|||
|
DQ 179 ACCAAACTGCGCTCACATTTAAATCTGAGCGCATCTCTTTGTTCCGAGGTTCTCACT 238
|||||
|||
|
Db 335 SerAspLeuAlaLeuLeuLeuLyrgerGIyAlaLeuProAlaProLeuLysValGIuIn 354
|||||
|||
|
DQ 239 GAAGAGACGATCTCTTCTGATCTTGGAAAAACAATGACACAGGCAATTATCTCAGA 298
::: |||
|||
|
Db 355 GIAsrThrValGIyAlaGIuLeuGIyAlaAspAlaValArgAlaGIyAlaIleSerThr 374
::: |||
|||
|
DQ 299 TGCTGTGGCTGGCAATGCTATGTTTGTGATGAGCGGTATTTAT-----AGATTTGA 352
|||||
|||
|
Db 375 LeuValAlaPheIleThrIleValValPheMetIleLeuSerTrpGIyLeuLeuPheGIy 394
|||||
|||
|
DQ 353 GGCGTCATGCGCTTGGAGACTGTTCTTCTGAATCTTTGCTTATCTGAGCACTCTACAG 412
|||
|||
|
Db 395 GIY---IleSerAlaIleAlaLeuIleIleAsnGIyMetLeuIleValAlaIleMetSer 413
|||
|||
|
DQ 413 TATTTGGATGGCCCACTGCTTGTCTGAGACTGCTGGGATGTTTGTCTATGGGAG 472
|||||
|||
|
Db 414 LeuThrGlAlaIleThrLeuThrLeuProGIyIleAlaGIyLeuIleLeuThrIleuAlaVal 433
|||||
|||
|
DQ 473 GGCGTAGATGCAAAATGTTCTTGTATTCGAAAGATCCGAGAGATTTTATTTGTTCTCA 532
|||||
|||
|
Db 434 AlaValAspAlaAsnValLeuIleTrpGIuArgMetArgAspAlaValArgAlaGIyLys 453
|||||
|||
|
DQ 533 AGCTTTAAAAATCTGTAGAAAAAGATATACCAGAGCTTTTGGACCATTTTATTTCT 592
|||||
|||
|
Db 454 SerProIIleuAlaIleAspAlaGIyPheSerArgAlaMetThrThrIleIleAspAla 473
|||||
|||
|
DQ 593 AACTTGACTTACAGATTTGGCTTCAGACATTTCTTTCTCTAGATACAGGCGCTATTA 652
|||||
|||
|

[illegible]

E83169
 C:Species: Pseudomonas aeruginosa (strain PA01)
 C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
 C:Accession: E83169
 R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Badian, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim, J.; Lory, S.; Olson, M.V.
 Nature 406, 959-964, 2000
 A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen
 A:Reference number: A82950; MUID:20437337; PMID:10984043
 A:Accession: E83169
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-306 <STO>
 A:Cross-references: GB:AE004799; GB:AE004091; NID:g9949981; PIDN:AAG07207.1; GSPDB:GN001
 A:Experimental source: strain PA01
 C:Genetics:
 C:Superfamily: Escherichia coli preprotein translocase chain secF

Alignment Scores:
 Pred. No.: 3,75e-24 Length: 306
 Score: 369.50 Matches: 91
 Percent Similarity: 51.69% Conservative: 77
 Best Local Similarity: 28.00% Mismatches: 102
 Query Match: 11.16% Indels: 55
 DB: 2 Gaps: 10

US-09-868-987-1 (1-1864) x E83169 (1-306)

QY 779 TTCTGGGGTAAGCATGATTTCTGAGAGGATGCAAAACTTTGGCTGTTCTGGA 838
 Db 10 PheMetGlyLeuArgAsn-----ValAlaPhe 18
 QY 839 AGTGTCTTCTTTAGTGTGGTCTCTCGGTTTGGAGCCTCGAATTCGTT----- 892
 Db 19 AlaValThrLeuIle-----LeuThrValIleAlaLeuGlySerIlePheThrLysGlyIle 37
 QY 893 ---TTGGGAATGGATTTAAAGAGGGGTATGCCCTTTACCTTTAATCCAAAGAGCATGGC 949
 Db 38 AsnPheGlyLeuAspPheThrGlyGlyThrLeuIleGluLeuThrTyrrGluInPro---- 56
 QY 950 ATCAGCATGTTGCTCAATCGTGGCAAGTGTGTCATAAATACAGGAGCTGGTCTT 1009
 Db 57 ---AlaAspLeuGlyLysValArgGlyGlnLeuVal-----GlyAla 69
 QY 1010 TCTTCTAGAGACTTCGCTATTCACACATTTGGATCTTCAGAAAGATCAAAATCTATTT 1069
 Db 70 GlyTyrrGluAspAlaValValGlnSerPheGlyAspAlaArgAspValLeuValArgMet 89
 QY 1070 AGTGATAAGCTTTTAAGCTATATCAAGCAGATACGAGCCTCTCTCTAAATTAACGATC 1129
 Db 90 ProSerGluAspProGluLeuGlyLysValAlaThrAlaLeuGlnAlaAsp--- 108
 QY 1130 ATGAGCTGGCGTTATTTGGGATTTGTTCTCAGAAACAGGCGCTAGATTTCTCTACGGAAC 1189
 Db 109 -----AlaGly 110
 QY 1190 TCTAAACGAACGCAAAATTTTGGTCAAGGTAGCAGCAAACTATCGAAGAAATCGGT 1249
 Db 111 AsnProAlaAsnLeuLysArgValGluTyrrValGlyProGlnValGlyGluGluLeuArg 130
 QY 1250 TATCAGCGACCATCGGCTTTTAGGAGCTTTGGCAATCATCTTGTCTATGTCAGTTTG 1309
 Db 131 AspGlnGlyLeuGlyMetLeuAlaLeuGlyIleLeuLeuTyrrValGlyPhe 150
 QY 1310 CGCTTGAATGGCAATATGCTTTCAGTCCGCTTAATTCATGACCTTTTGGCT 1369
 Db 151 ArgPheGlnTrpLysPheAlaLeuGlyAlaIleLeuSerLeuValHisAspAlaIleIle 170
 QY 1370 ACCTGTGAGTCTTGTATTAGCACATTCTTTTGGAGAAATTCAAATAGATTTGCA 1429
 Db 171 ValMetGlyValLeu-----SerPhePhe-----GlnValThrPheAspLeuThr 185
 QY 1430 GCCATTGCTGTTTAATGACTGTATCGGGTATTTCATTAACAATCTGATCATTTT 1489
 Db 186 ValLeuAlaValLeuAlaValValGlyTyrrSerLeuAsnAspThrIleValIlePhe 205
 QY 1490 GATCGTATTGCTGAAGAT-----CGCAACGGAACTGTTTACCCCTTCATCAT 1537
 Db 206 AspArgValArgGluAsnPheArgValLeuArgLysAlaAspLeuValGluAsnLeuAsn 225
 QY 1538 GTTTTGTAAATGATGCCCTTCAAAAGACGTTCCAGCGCAGGTAATGACACACAGCTACA 1597
 Db 226 Ile-----SerThrSerGlnThrLeuLeuArgThrIleAlaThrSerValSer 241
 QY 1598 ACTCTATCAGTTTTGTAAATGCTTTTATATAGCGCGCTCTCTGCTTTTAAATTTGCA 1657
 Db 242 ThrLeuLeuAlaIleAlaAlaLeuLeuPheGlyGlyAspAsnLeuPheGlyPheSer 261
 QY 1658 TTATATGACCATGAGGATCTTCTAGGAACCTTATCGTCTCTTATATATGACCACT 1717
 Db 262 IleAlaLeuPheValGlyValMetAlaGlyThrTyrrSerSerIleTyrrIleAlaAsnVal 281
 QY 1718 CTGTTGTTGTTTATG 1732
 Db 282 ValLeuIleTrpLeu 286

RESULT 34
 F82429
 protein-export membrane protein SecD VCA0693 [imported] - Vibrio cholerae (strain N16961)
 C:Species: Vibrio cholerae
 C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 14-Sep-2001
 C:Accession: F82429
 R:Heidelberger, J.P.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Dodson, R.J.; Chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bae, S.; Qin, H.; Dragoi, I.; Sellers, P. I., R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
 Nature 406, 477-483, 2000
 A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
 A:Reference number: A82035; MUID:20406833; PMID:10952301
 A:Accession: F82429
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-612 <HEI>
 A:Cross-references: GB:AE004398; GB:AE003853; NID:g9658102; PIDN:AAF96592.1; GSPDB:GN001:
 A:Experimental source: serogroup O1; strain N16961; biotype El Tor
 C:Genetics:
 A:Gene: VCA0693
 A:Map position: 2
 C:Superfamily: protein export membrane protein secD

Alignment Scores:
 Pred. No.: 5,79e-24 Length: 612
 Score: 368.00 Matches: 79
 Percent Similarity: 59.17% Conservative: 63
 Best Local Similarity: 32.92% Mismatches: 94
 Query Match: 11.11% Indels: 4
 DB: 2 Gaps: 3

US-09-868-987-1 (1-1864) x F82429 (1-612)

QY 38 ACTGCTTAATCGCAATATCTGCAAAACCGTGGATGGCTGATGCTGATGATGACGGT 97
 Db 370 ThrValTyrrArgGluTyrrLysThrAsnAla-----ArgGlyGluThrValArgSerGlu 387
 QY 98 TATATGCTCAGCAGCCCTATTTTAAAGCTCCCATTTGAAATCATGCCAGTGTCTCAGG 157
 Db 388 LysValIleSerValAlaThrIleGlnSerGlnLeuGlySerGlnPheArgIleThrGly 407
 QY 158 AAATTTACCCACCGTGAAGTGAGCAACTCGCTCAGATTTAAATCTGGAGCATGTCT 217
 Db 408 AlaGlySerMetGluGluAlaGlnGlnLeuAlaLeuLeuArgAlaGlySerLeuThr 427
 QY 218 TTTGTTCCCGAGTTCTCAGTGAAGACGATCTTCTGATCTTGGGAAAAAACAATGT 277

US-09-868-987-1 (1-1864) x AH0552 (1-323)

```
Qy 794 CATGATTTCTTTGAGAGGATGCAAAAACCTTTGGCT-----GTTCTGGAAGTGTTTT 847
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 17 TyrAspPheMetArgTTP-----AspPheTrpAlaPheGlyIleSerGlyLeuLeu 34

Qy 848 CTTTATAGTTCGGTTCCTCGGTTTGGAGCCTGGAATTCGGTTTGGGAATGATTTT 907
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 35 IleAlaIleValIleMetGlyValArgGlyPheAsn-----TrpGlyLeuAspPhe 52

Qy 908 AAAGAGGGTATGCTTTACCTTTAATCCAAAGAGCATGCGATCAGCGATGTGGCTCAA 967
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 53 ThrGlyGlyThrValIleGluIleThrLeuGluIysProAlaGluMetAspVal----- 70

Qy 968 ATGGCTGCAAAAGTTGTCATAAACTACAGGAAGCTGCTCTTCTTCTAGAGACTTCGT 1027
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 71 MetArg-----GluAlaLeuGlnIysAlaGlyTyr-----GluGluProGln 84

Qy 1028 ATTCAAACATTTGGATCTTCAGAAAGATCAAAATCTATTTTAGTCATAAGCTTTAAGC 1087
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 85 LeuGlnAsnPheGlySerSerHisAspIleMetValArgMetProThrGluGlyGlu 104

Qy 1088 TATACTAAGCAGATGAGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1147
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 105 ThrGlyGlyGlnValLeuGlySerIysVal----- 114

Qy 1148 GGSATTTGTTGAGAAACAGCGCTAGATTTCTCTACGGAATCTCTAAACGAAACGCAAA 1207
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 115 ---ValThrIleIleAsnGluAlaThr-----AsnGlnAsnAlaValIys 129

Qy 1208 TTTTGGTCAAAGGTGAAGCAGCAACTATCGAAGAAATGCGTTATCAGCGCAACATCGG 1267
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 130 ArgIleGluPheValGlyProSerValGlyAlaAspLeuAlaGlnThrGlyAlaMetAla 149

Qy 1268 CTTTATAGGAGCTTTGGCAATCATCTGCTCTATGTGAGCTTTGCGCTTTGAATGGCAAT 1327
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 150 LeuLeuValAlaLeuIleSerIleLeuValIyrValGlyPheArgPheGluTrpArgLeu 169

Qy 1328 CTTTACGTGCGGTATGCGCTTTAATTCATGACCTTTTGGCTACCTGTGCGCTTTGTT 1387
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 170 AlaAlaGlyValValIleAlaLeuAlaHisAspValIleIleThrLeuGlyIleLeuSer 189

Qy 1388 ATAGCAGATTTCTTTTGAAGAAATTCAAATAGATTTGCAAGCCATGCTGCTTTAATG 1447
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 190 LeuPheHis-----IleGluIleAspLeuThrIleValAlaSerLeuMet 204

Qy 1448 ACTGTATTGGGTATTTCATTAACAATACTTTGATCATTTTGTATTCGTTCGTCAGAT 1507
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 205 SerValIleGlyTyrSerLeuAsnAspSerIleValSerAspArgIleArgGluAsn 224

Qy 1508 ---CGCCAAGCGAACCTTTTACCCTATGATGTTTATGTTAATGATGCGCTTCAAAAG 1564
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 225 PheArgIysIleArgArgGlyThrProTyrGluIlePhe---AsnValSerLeuThrGln 243

Qy 1565 ACGTTCACCGCAGCGTAATGACACACACTCAACTATCATGTTTCTTAAATGCTTTTG 1624
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 244 ThrLeuHisArgThrLeuIleThrSerGlyThrLeuValValIleLeuMetLeuTyr 263

Qy 1625 TTTATAGCGCTCTCTCTCTTTAATTTTCATTTATTATGACCATAGGATCTTCTTA 1684
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 264 LeuPheGlyGlyProValLeuGluGlyPheSerLeuThrMetLeuIleGlyValSerIle 283

Qy 1685 GGAACCTTTATCGTCTCTTTATATTGACCACTCTGTTGTTG 1726
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 284 GlyThrAlaSerSerIleTyrValAlaSerAlaLeuAlaLeu 297
```

RESULT 37

JQ0697

preprotein translocase chain secY [validated] - Escherichia coli (strain K-12)

N;Alternate names: protein-export membrane protein secY; secretion protein secF

C;Species: Escherichia coli

C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 01-Mar-2002

C;Accession: JQ0697; S12302; A64770

R;Gardel, C.; Johnson, K.; Jacq, A.; Beckwith, J.

EMBO J. 9, 3209-3216, 1990

A;Title: The secD locus of E.coli codes for two membrane proteins required for protein e

A;Reference number: JQ0693; MUID:91006014; PMID:2170107

A;Accession: JQ0697

A;Molecule type: DNA

A;Residues: 1-323 <GAR>

A;Cross-references: GB:X56175; NID:g42929; PIDN:CAA39635.1; PID:g581231

A;Experimental source: strain K-12

R;Gardel, C.; Johnson, K.; Jacq, A.; Beckwith, J.

EMBO J. 9, 4205-4206, 1990

A;Reference number: S12298; MUID:91065354; PMID:2249673

A;Contents: erratum

A;Accession: S12302

A;Molecule type: DNA

A;Residues: 1-323 <GA2>

A;Cross-references: EMBL:X56175; NID:g42929; PID:g581231

R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Col

A.; Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A;Title: The complete genome sequence of Escherichia coli K-12.

A;Reference number: A64720; MUID:97426617; PMID:9278503

A;Accession: A64770

A;Status: nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-323 <BLAT>

A;Cross-references: GB:AE000147; GB:U00096; NID:g1786503; PIDN:AAC73512.1; PID:g1786510;

A;Experimental source: strain K-12, substrain MG1655

C;Comment: Preprotein translocase contains a membrane-embedded trimeric complex of SecY, heterotrimeric complex. These two trimeric complexes are associated to form SecYEGDFYajC

C;Genetics:

A;Gene: secF

A;Start codon: GTG

C;Complex: heterohexamer; chains secY (PIR:QOECY), secE (PIR:VXECSE), secG (PIR:S40402),

C;Function:

A;Description: the secD protein is a transmembrane component of the protein export complex

C;Superfamily: Escherichia coli preprotein translocase chain secF

C;Keywords: inner membrane; protein export; transmembrane protein

F;27-43/Domain: transmembrane #status predicted <TM1>

F;147-163/Domain: transmembrane #status predicted <TM2>

F;172-188/Domain: transmembrane #status predicted <TM3>

F;195-211/Domain: transmembrane #status predicted <TM4>

F;249-265/Domain: transmembrane #status predicted <TM5>

F;275-291/Domain: transmembrane #status predicted <TM6>

Alignment Scores:

Pred. No.:	1-15e-23	Length:	323
Score:	364.00	Matches:	102
Percent Similarity:	49.52%	Conservative:	53
Best Local Similarity:	32.59%	Mismatches:	124
Query Match:	10.99%	Indels:	34
DB:	1	Gaps:	9

US-09-868-987-1 (1-1864) x JQ0697 (1-323)

Qy 794 CATGATTTCTTTGAGAGGATGCAAAAACCTTTGGCTGTCTTCTGGAAGTGTTTTCTTTTA 853

:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

Db 17 TyrAspPheMetArgTTPAspTyrTrpAlaPheGlyIleSerGlyLeuLeuIleAla 36

:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

Qy 854 GGTTCGCTGCTCTCGGTTTGGAGCCTGGAATTCGGTTTGGGAATGGATTTTAAAGGA 913

:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

Db 37 AlaIleValIleMetGlyValArgGlyPheAsn-----TrpGlyLeuAspPheThrGly 54

:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

Qy 914 GGSPATGCCITTTACCTTTAATCCAAAAGAGCATGGCATCAGCGATGTTCCTCAATGGGT 973

:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

Db 55 GlyThrValIleGluIleThrLeuGluIysProAlaGluIleAspVal-----MetArg 72

:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

Qy 974 GGCAAGTTGTGCATAAATACAGAAAGCTGGTCTTTCTCTAGAGACTTCGCTATTCAA 1033

:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

Db 73 -----AspAlaLeuGlnIysAlaGlyPhe-----GluGluProMetLeuGln 86

:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

Qy 1034 ACATTGGATCTTCAGAAAAGATCAAAATCTTTATTAGTAGATAAAGCTTTAAGCTATACT 1093

:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

Db 87 AsnPhgIySerSerHisAspIleMetValArgMetProProAlaGluGlyValThrGly 106

Qy 1094 AAGCAGATACGAGCCTCT---CTCCTAAATTAACGATCATGAGCTGGCGTTATGTGGG 1150

Db 107 GlyIleuValLeuGlySerGlnValLeuIleuValIleAsnGluSer----- 121

Qy 1151 ATTGTTGTTCAGAAACAGCGCTGATTTCTCTACGGAACTCTAAACGAAACGAAATTT 1210

Db 122 -----ThrAsnGlnAsnAlaIleValysArg 130

Qy 1211 TGGTCMAAGTAAAGCAGCAACTATCGAAGAAATCGCTTATCAGCGACCATCGGGCTT 1270

Db 131 IleGluPheValGlyProSerValAlaGlyAlaAspLeuAlaGlnThrGlyAlaMetAlaLeu 150

Qy 1271 TTAGAGAGCTTTGGCAATCATCTTGTCTTATGTGAGATTGGCTGTGAATGCAATATGCT 1330

Db 151 MetAlaIleLeuLeuSerIleLeuValIleValIlePheArgPheGluTrpArgLeuAla 170

Qy 1331 TTCAGTCCGCTATGCGCTTAACTCATGACCTTTGGCTACCTGGACGCTGTGTTATA 1390

Db 171 AlaGlyValAlaIleAlaLeuAlaHisAspValIleIleThrLeuGlyIleLeuSerLeu 190

Qy 1391 GCACATTTCTTTTGAAGAAATTCCAATAGATTGGCAAGCATTGCTGCTTAAATGACT 1450

Db 191 PheHis-----IleGluIleAspLeuThrIleValAlaSerLeuMetSer 205

Qy 1451 GTATTGGGTTATTCATTAACAATACTTGATCATTTTGTATGCTGATTCGTAAGAT--- 1507

Db 206 ValIleGlyIleTyrSerLeuAsnAspSerIleValIleValSerAspArgIleArgIleAsn 225

Qy 1508 CGCCCAAGCAACCTGTTTACCCCTATGCATGTTTAAATGATGAGCCCTGCAAAAGCG 1567

Db 226 ArgIleValIleArgArgGlyThrProTyrGluIlePhe--AsnValSerLeuThrGlnThr 244

Qy 1568 TTCAGCGCGCAGGTATGACCAACAGCTACACTCTATCAGTTTGTATGCTTGTGTT 1627

Db 245 LeuHisArgThrLeuIleThrSerGlyThrThrLeuMetValIleLeuMetLeuTyrLeu 264

Qy 1628 ATAGCGGGCTCCTGCTGCTTAAATTTGGCATTTATTAAGCAATAGAGATCTTTCAGCA 1687

Db 265 PheGlyGlyProValLeuGluGlyPheSerLeuThrMetLeuIleGlyAlaSerIleGly 284

Qy 1688 ACTTATCTCTCTTTATATGTGACACCACTCTGTTGTTG 1726

Db 285 ThrAlaSerSerIleTyrValAlaSerAlaLeuAlaLeu 297

RESULT 38

D90686

protein-export membrane protein secf [imported] - Escherichia coli (strain O157:H7, sub

C/Species: Escherichia coli

C/Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001

C/Accession: D90686

R/Hayashi, T.; Makino, K.; Onishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.

gaganara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hataori, M.; Shinagawa, H.

DNA Res. 8, 11-22, 2001

A/Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gene

A/Reference number: A39629; MUID:21156231; PMID:11258796

A/Accession: D90686

A/Status: preliminary

A/Molecule type: DNA

A/Restrictions: 1-333 <HAY>

A/Cross-References: GB:BA000007, PIDN:BA833883.1, PID:gl3359917, GSPDB:GN00154

A/Experimental source: strain O157:H7, substrain RIMD 0509952

C/Genetics:

A/Gene: EC90460

C/Superfamily: Escherichia coli prepore protein translocase chain secf

Alignment Scores:

Pred. No.: 1,156-23 Length: 323

Score: 364.00 Matches: 102

Percent Similarity: 49.52% Conservative: 53

Best Local Similarity: 32.58% Mismatches: 124

Query Match: 10.99% Indels: 34

[illegible]

C;Accession: H85536

R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimailanta, B.; Potamoudis, K.; Apodaca,
Nature 409, 529-533, 2001

A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.

A;Reference number: A85480; MUID:21074935; PMID:11206551

A;Accession: H85536

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-323 <STO>

A;Cross-references: GB:AE0051174; NID:g12513259; PIDN:AAG54756.1; GSPDB:GN00145; UWGP:Z05

A;Experimental source: strain O157:H7, substrain EDL933

C;Genetics:

A;Gene: secF

C;Superfamily: Escherichia coli preprotein translocase chain secF

Alignment Scores:

Pred. No.:	1.15e-23	Length:	323
Score:	364.00	Matches:	102
Percent Similarity:	49.52%	Conservative:	53
Best Local Similarity:	32.59%	Mismatches:	124
Query Match:	10.99%	Indels:	34
DB:	2	Gaps:	9

US-09-868-987-1 (1-1864) x H85536 (1-323)

QY 794 CATGATTTCTGAGAGGATGCAAAAACCTTTGGCTGTTCTGGAAGTGTTCCTTTTA 853

Db 17 TyrAspPheMetArgTyrTrpAlaPheGlyIleSerGlyLeuLeuIleAla 36

QY 854 GGTTCCTGCTCTCGGTTGGAGCTCGGATTCGTTTGGCAATGGATTTAAAGGA 913

Db 37 AlaIleValIleMetGlyValArgGlyPheAsn-----TrpGlyLeuAspPheThrGly 54

QY 914 GGGTATGCTTTACCTTTAATCAAAAGATCGCATCGCGATGTTCCTCAATCGT 973

Db 55 GlyThrValIleGluIleThrLeuGluIlePsrProAlaGluIleAspVal-----MetArg 72

QY 974 GGCAAGTTGTGCATAACTACAGAGAGCTGGTCTTTCTCTAGAGACTTCGCTATTCAA 1033

Db 73 -----AspAlaLeuGlnLysAlaGlyPhe-----GluGluProMetLeuGln 86

QY 1034 ACATTGGATCTTCAGAAAGATCAAAATCTATTTTAGTGATAAAGCTTTAAGCTACT 1093

Db 87 AsnPheGlySerSerHisAspIleMetValArgMetProAlaGluGlyGluThrGly 106

QY 1094 AAGCAGATACGAGCTCT---CTCCTAAATTAACGATCATGAGCTGGCTTATTGGG 1150

Db 107 GlyGlnValLeuGlySerGlnValIleLysValIleAsnGluSer----- 121

QY 1151 ATTGTTGTCAAAACAGGCTTAGATTTCTTCAGGAAACTCTAAACGAAATTT 1210

Db 122 -----ThrAsnGlnAsnAlaAlaValIleArg 130

QY 1211 TGGTCAAGGTAAACGAAACTATCGAAGAAATCGGTATACAGGACCATCGGCTT 1270

Db 131 IleGluPheValGlyProSerValGlyAlaAspLeuAlaGlnThrGlyAlaMetAlaLeu 150

QY 1271 TTAGAGCTTTGGCAATCATCTCTATGTAGTTGGCTTTGATGCAATATGCT 1330

Db 151 MetAlaLeuLeuSerIleLeuValIleValIleThrLeuGlyIleLeuSerLeu 190

QY 1331 TTCAGTCCGCTATCGCTTTAAATTCATGACCTTTTGGCTACTGCTGTCGCTCTTTATA 1390

Db 171 AlaGlyValValIleAlaLeuAlaHisAspValIleIleThrLeuGlyIleLeuSerLeu 190

QY 1391 GCACATTTCTTTTGAAGAAATTCATATAGATTGCAAGCCATGTGCTTTAATGACT 1450

Db 191 PheHis-----IleGluIleAspLeuThrIleValAlaSerLeuMetSer 205

QY 1451 GTATTGGGTATTTCATTAACAATTCATTGATCTTTTGTATCTGATCTGAGAT---- 1507

Db 206 ValIleGlyIleSerLeuAsnAspSerIleValValSerAspArgIleArgGluAsnPhe 225

QY 1508 CGCAACGCAACCTGTTTACCCCTATGTCATGTTTATTAGTATGATGCCCTTCAAAAGACG 1567

Db 226 ArgIysIleArgArgGlyThrProTyrGluIlePhe---AsnValSerLeuThrGlnThr 244

QY 1568 TTCAGCGCGCAGGTAAATGACCAACAGCTACACTCTATCAGTCTTTGTTAAAGTCTTTTCT 1627

Db 245 LeuHisArgThrLeuIleThrSerGlyThrThrLeuMetValIleLeuMetLeuLeu 264

QY 1628 ATAGGCGGCTCCTCTGCTTTTAAATTTTATGATACCATAGGATTTCTTCTAGGA 1687

Db 265 PheGlyGlyProValLeuGluGlyPheSerLeuThrMetLeuIleGlyValSerIleGly 284

QY 1688 ACTTATCGTCTCTTATATGTCACCACTCTCTGTTGTTG 1726

Db 285 ThrAlaSerSerIleTyrValAlaSerAlaLeuAlaLeu 297

RESULT 40

C81926

probable protein-export membrane protein NMA0812 [imported] - Neisseria meningitidis (str)

C;Species: Neisseria meningitidis

C;Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 14-Sep-2001

C;Accession: C81926

R;Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morell

; Holroyd, S.; Jagers, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,

Nature 404, 502-506, 2000

A;Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.

A;Reference number: A81775; MUID:2022556; PMID:10761919

A;Accession: C81926

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-618 <PAR>

A;Cross-references: GB:AL162754; GB:AL157959; NID:g7379424; PIDN:CAB84094.1; PID:g737953

A;Experimental source: serogroup A, strain Z2491

C;Genetics:

A;Gene: secD; NMA0812

C;Superfamily: protein export membrane protein secD

Alignment Scores:

Pred. No.:	1.3e-23	Length:	618
Score:	364.00	Matches:	80
Percent Similarity:	55.95%	Conservative:	61
Best Local Similarity:	31.75%	Mismatches:	105
Query Match:	10.99%	Indels:	6
DB:	2	Gaps:	3

US-09-868-987-1 (1-1864) x C81926 (1-618)

QY 23 GAGGGATACGCGCACTGCTAATGGACAATAT---TCTGCAAAACCGTGGATGGCGTATG 79

Db 348 AspSerAlaGlyGlySerIlePheGlyGluLeuThrAlaAlaAsnValGlyLysArgMet 367

QY 80 GCTGTAGTGATT-----GACGGTTATATGTCACGACCCCTATTTTAAACGTC 127

Db 368 AlaMetValLeuIleAspGingIlyLysSerGluValThrAlaProValIleArgThr 387

QY 128 CCATTGAAA---AATCATGCCAGTCTCAGGGAATTTTACCCACCGCTAGTGCAAA 184

Db 388 AlaIleThrGlyArgValGluIleSerGlySerMetThrThrAlaGluAlaAsnAsp 407

QY 185 CTCGCTCAGATTTAAATCTGGAGCGATGCTTTTGTCTCCGAGGTTCTCAGTGAAGAG 244

Db 408 ThrSerLeuLeuArgAlaGlySerLeuAlaProMetGlnIleValGluArg 427

QY 245 ACATCTCTCTCATCTTGGGAAAAAACAATGTACACAGGCATTATCTCAGCATGCTGT 304

Db 428 ThrIleGlyProSerLeuGlyLysGluAsnIleGluLysGlyPheHisSerThrLeuTrp 447

QY 305 GGCTGGCAATGCTATTGTTTGTATGAGCGTATATTATAGATTTTGGAGCGCTCATCGCT 364

Db 448 GlyPheAlaIleValAlaAlaPheMetValTyrTyrArgLeuMetGlyPhePheSer 467

QY 365 TCGGAGAGCTGCTTCTTGAAATCTTTTCTTATCTGGGACGCTCTACAGTATTTGGATCGC 424

|||||:||||| ||| :|||:||||| |||
Db 12 AspPheMetArgTrpSerLysPheAlaPheAlaLeuSerLeuValMetIleAlaAlaSer 31
Qy 842 GTTTTCTTTTAGGTGCTCTCGGTTTGGAGCTGGAATTCGTTTGGGAATG 901
:||||| ||| :|||:||||| |||
Db 32 IlePheThrLeuSerThrLysTrpLeuAsn-----Trp-----GlyLeu 44
Qy 902 GATTTAAAGAGAGGTATGCCTTTACCTTTTAATCCAAAGAGCATGGCATCAGCGATGTT 961
||||| ||| :|||:||||| |||
Db 45 AspPheThrGlyThrLeuIle-----GluValGlyPheGluGlnPro 59
Qy 962 GCTCAATGCGTGGCAAGTTGTGCATAAACTACAGGAAGCTGCTTCTTCTTAGAGAC 1021
||||| ||| :|||:||||| |||
Db 60 AlaAsnLeuGlu---GlnIleArgSerAlaLeuGluAlaLysGlyPheGly-----Asp 76
Qy 1022 TTCGGTATTCAACCATTTGGATCTTTCAGAAAGATCAAAATCTAT----- 1066
:||||| ||| :|||:||||| |||
Db 77 AlaThrValGlnAsnPheGlySerAlaArgAspValMetValArgLeuArgProArgAsp 96
Qy 1067 ---TTTAGTGATAAGCTTTAAGCTATACTAAGCAGATACGAGCTCTCTCTAAATTA 1123
:||||| ||| :|||:||||| |||
Db 97 AspMetAlaGlyGluAlaLeuGly-----AsnGlnIleLeuAlaIleLysGluGly 114
Qy 1124 ACGATCATGACCTGGCGTTATGTGGGATTCTTGTCAAAACAGCGCTAGATTTCTCTAC 1183
||||| ||| :|||:||||| |||
Db 115 Thr-----GlyGlyAsnValGluMetArgArgIleGluPhe----- 126
Qy 1184 GGAACCTCTAAACGAAACCAAAATTTTGGTCAAAAGTAAGCAGCAAACTATCGAAGAAA 1243
:||||| ||| :|||:||||| |||
Db 127 -----ValGlyProAsnValGlyAspGlu 134
Qy 1244 ATGCGTATACAGCGACCATCGGGCTTTAGAGAGCTTTGGCAATCATCTTCTCTATGTG 1303
:||||| ||| :|||:||||| |||
Db 135 LeuThrGluAlaGlyGlyLeuAlaIleLeuValSerLeuLeuCysIleLeuLeuTyrVal 154
Qy 1304 ACTTGGCGCTTTGAATGCAATATGCTTTCAGTCCGATGCGCTTTAATTCATGACCTT 1363
:||||| ||| :|||:||||| |||
Db 155 SerValArgPheGluTrpArgLeuAlaAlaGlyAlaValLeuAlaLeuAlaHisAspVal 174
Qy 1364 TTGGCTACCTGCGACGTCTGTTTATAGCACATTTCTTTTGMAGAAATTCAAATAGAT 1423
:||||| ||| :|||:||||| |||
Db 175 IleIleThrLeuGlyIle-----PheSerIleLeuGlnIleGluValAsp 189
Qy 1424 TTGCAAGCCATTGGTCTTTAATGACTATGATGGGATTCATTAAACAATCTTTGATC 1483
:||||| ||| :|||:||||| |||
Db 190 LeuThrIleValAlaAlaLeuLeuThrValValGlyTyrSerLeuAsnAspThrIleVal 209
Qy 1484 ATTTTGTGCTATTCGTGAAGAT---CGCCAAGCGAACCTGTTTACCCCTATGATGTT 1540
:||||| ||| :|||:||||| |||
Db 210 ValPheAspArgIleArgGluAsnPheArgLysMetArgLysGluGluProAlaGluIle 229
Qy 1541 TTAGTTAATGATGCCCTTCAAAAGACGTTACAGCCGCGTAATGACAAACAGCTACAACT 1600
:||||| ||| :|||:||||| |||
Db 230 Met---AsnSerSerIleThrGlnThrLeuSerArgThrLeuIleThrSerGlyThrThr 248
Qy 1601 CTATCAGTTTTTGTAAATGCTTTTGTGTTATAGCGGCTCTCTGCTCTTAATTTTGCATTT 1660
:||||| ||| :|||:||||| |||
Db 249 LeuPheValIleAlaLeuPheThrGlnGlyAlaMetIleHisGlyPheAlaLeu 268
Qy 1661 ATTTAGCAATAGGATTCCTCTAGGAACCTTTATCGTCTCTTTATATGACCACTCTG 1720
:||||| ||| :|||:||||| |||
Db 269 AlaLeuLeuLeuGlyIleThrValGlyThrTyrSerSerIleTyrValAlaSerAlaLeu 288
Qy 1721 TTGTTG 1726
Db 289 AlaLeu 290

RESULT 43
H64769

preprotein translocase chain secD [validated] - Escherichia coli (strain K-12)
N;Alternate names: protein-export membrane protein secD; secretion protein secD
C;Species: Escherichia coli
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 01-Mar-2002

C;Accession: H64769; JQ0696; S12301
R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Col
.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A;Title: The complete genome sequence of Escherichia coli K-12.
A;Reference number: A64720; MUID:97426617; PMID:9278503
A;Accession: H64769
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-615 <BLAT>
A;Cross-references: GB:AE000147; GB:U00096; NID:G1786603; PIDN:AAC73511.1; PID:G1786609;
R;Gardel, C.; Johnson, K.; Jacq, A.; Beckwith, J.
EMBO J. 9, 3209-3216, 1990
A;Title: The secD locus of E.coli codes for two membrane proteins required for protein e;
A;Reference number: JQ0693; MUID:91006014; PMID:2170107
A;Accession: JQ0696
A;Molecule type: DNA
A;Residues: 1-77, 'S', 79-154, 'A', 156-615 <GAR>
A;Cross-references: GB:X56175; NID:G42929; PIDN:CAA39634.1; PID:G581230
R;Gardel, C.; Johnson, K.; Jacq, A.; Beckwith, J.
EMBO J. 9, 4205-4206, 1990
A;Reference number: S12298; MUID:91065354; PMID:2249673
A;Contents: erratum
A;Accession: S12301
A;Molecule type: DNA
A;Residues: 1-77, 'S', 79-154, 'A', 156-615 <GA2>
A;Cross-references: EMBL:X56175; NID:G42929; PID:G581230
R;Pogliano, K.J.; Beckwith, J.
J. Bacteriol. 176, 804-814, 1994
A;Title: Genetic and molecular characterization of the Escherichia coli secD operon and i
A;Reference number: A36969; MUID:94131960; PMID:7507921
C;Contents: annotation; membrane topology
C;Genetics:
A;Gene: secD
A;Start codon: GTG
C;Complex: heterohexamer; chains secY (PIR:QOEGSY), secE (PIR:VKESE), secG (PIR:S40402),
ded trimeric complex of secY, secE and secG and the peripheral secA protein; the protein
associated to form SecYEGDyajC, the hexameric integral membrane domain of the pre-protein
C;Function:
A;Description: the secD protein is a transmembrane component of the protein export comple
C;Superfamily: protein export membrane protein secD
C;Keywords: inner membrane; protein export; transmembrane protein
F;10-30/Domain: transmembrane #status predicted <TM1>
F;31-455/Domain: periplasmic #status predicted <PP1>
F;456-472/Domain: transmembrane #status predicted <TM2>
F;477-497/Domain: transmembrane #status predicted <TM3>
F;498-501/Domain: periplasmic #status predicted <PP2>
F;502-518/Domain: transmembrane #status predicted <TM4>
F;564-580/Domain: transmembrane #status predicted <TM5>
F;581-585/Domain: periplasmic #status predicted <PP3>
F;586-605/Domain: transmembrane #status predicted <TM6>
Alignment Scores:
Pred. No.: 3,23e-23 Length: 615
Score: 359.50 Matches: 80
Percent Similarity: 62.39% Conservative: 61
Best Local Similarity: 35.40% Mismatches: 82
Query Match: 10.85% Indels: 3
DB: 1 Gaps: 2
US-09-868-987-1 (1-1864) x H64769 (1-615)
Qy 44 AATGACAATATTCGCAACCGTGGATGGCTATGGCTAGTGATTCAGCGTTATATG 103
:||||| ||| :|||:||||| |||
Db 375 SerGlyLysLysAspAlaAsn-----GlyArgAlaValLeuValLysGlnGluVal 392
Qy 104 GTCAGCAGCCCTATTTTAAACGTCCTCCATTCGCAATCATGCCAGTGTCTCAGGGAATTT 163
:||||| ||| :|||:||||| |||
Db 393 IleAsnIleAlaAsnIleGlnSerArgLeuGlyAsnSerPheArgIleThrGlyIleAsn 412
Qy 164 ACCCACCGTGAATGACCAAACTCGCTCAGATTTAAATCTCGAGCGATGCTCTTTGTT 223
:||||| ||| :|||:||||| |||

```

Db 413 AsnProhansgluAlaArgGlnLeuSerLeuLeuArgAlaGlyAlaLeuIleAlaPro 432
Qy 224 CCCAGAGTTCTCAAGTGAAGACATCTCTTCTGATCTTGGAAAAACAATGACAA 283
Db 433 IleglnIleValGlnGluArgThrIleGlyProThrLeuGlyMetGlnAsnIleGln 452
Qy 284 GGCATTATCTCAGCATGCTGGCTGGCAATGCTTATGTTTGGAGGGATATATAT 343
Db 453 GilyeuGluAlaCysLeuAlaGlyLeuLeuValSerIleLeuPheMetIleIlePheTyr 472
Qy 344 AGATTGGAGGCGTCATCGCTTGGAGCTGTCTTCTGAACTTTTCTTATCTGGCA 403
Db 473 LysIysPheGlyLeuIleAlaThrSerAlaLeuIleAlaAsnLeuIleAlaGly 492
Qy 404 GCTCTACAGTATTG--GATGCCCACTCACTTGTACAGACTCGCTGGATTGTTCTT 460
Db 493 IleMetSerLeuLeuProGlyAlaThrLeuSerMetProGlyIleAlaGlyIleValLeu 512
Qy 461 GCTATGGGAGTGGCCGTGATGCAATGTTCTGTATTCGAAGAATCGAGAGGATTT 520
Db 513 ThrLeuAlaValAlaValAspAlaAsnValLeuIleAsnGluArgIleLysGlnGluLeu 532
Qy 521 TTATTGTCCTCAAGCTTAAATAATCTGTAGAAAAGATATACCAAGCTTTGGAGCC 580
Db 533 SerAsnGlyArgThrValGlnGlnAlaIleAspGluGlyTyrArgGlyAlaPheSerSer 552
Qy 581 ATTTTGTACTTACTGACTACGATATGGCTTCAGCACTTCTTCTTCTTACATACA 640
Db 553 IlePheAspAlaAsnIleThrThrLeuIleLysValIleIleLeuValAlaGlyThr 572
Qy 641 GGGCCTATTAAAGGTTGCTTGTGACATGATTAGGAATTTCTCTCAAGTTTACG 700
Db 573 GlyAlaIleLysGlyPheAlaIleThrThrGlyIleGlyValAlaThrSerMetPheThr 592
Qy 701 GCTCTTTCATGACTTAA 718
Db 593 AlaIleValGlyThrArg 598

RESULT 44
C90686
protein-export membrane protein secD [imported] - Escherichia coli (strain O157:H7, sub
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001
C:Accession: C90686
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.;
gasawara, N.; Yasunaga, T.; Kunara, S.; Shiba, T.; Hatтори, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gene
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: C90686
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-615 <HY>
A:Cross-references: GB:BA000007; PIDN:BA833882.1; PID:q13359916; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain R1MD 0509552
C:Genetics:
A:Gene: ECs0459
C:Superfamily: protein export membrane protein secD

Alignment Scores:
Pred. No.: 3,23e-23 Length: 615
Score: 359.50 Matches: 80
Percent Similarity: 62.39% Conservative: 61
Best Local Similarity: 35.40% Mismatches: 82
Query Match: 10.85% Indels: 3
DB: 2 Gaps: 2

US-09-868-987-1 (1-1864) x C90686 (1-615)
Qy 44 AATGACATATTTCTGCAACCGTGTGATGCTGATGATGATGATGATGATGATG 103
Db 375 SerGlyLysLysAspAlaAsn-----GlyArgAlaValLeuValLysGlnGlnGluVal 392

```

```

Qy 104 GTTCAGACCCCTATTAAAGTCCATGAAATATCATGACATGCTGACGGGAATTT 163
Db 393 IleAsnIleAlaAsnIleGlnSerArgLeuGlyAsnSerPheArgIleTrgIleAsn 412
Qy 164 ACCACCGTGAAGTGAACAACTCGCTCAGATTTAAATCTGAGGAGATCTTTGTT 223
Db 413 AsnProhansgluAlaArgGlnLeuSerLeuLeuArgAlaGlyAlaLeuIleAlaPro 432
Qy 224 CCCAGAGTTCTCAAGTGAAGACATCTCTTCTGATCTTGGAAAAACAATGACAA 283
Db 453 GilyeuGluAlaCysLeuAlaGlyLeuLeuValSerIleLeuPheMetIleIlePheTyr 472
Qy 344 AGATTGGAGGCGTCATCGCTTGGAGCTGTCTTCTGAACTTTTCTTATCTGGCA 403
Db 473 LysIysPheGlyLeuIleAlaThrSerAlaLeuIleAlaAsnLeuIleAlaGly 492
Qy 404 GCTCTACAGTATTG--GATGCCCACTCACTTGTACAGACTCGCTGGATTGTTCTT 460
Db 493 IleMetSerLeuLeuProGlyAlaThrLeuSerMetProGlyIleAlaGlyIleValLeu 512
Qy 461 GCTATGGGAGTGGCCGTGATGCAATGTTCTGTATTCGAAGAATCGAGAGGATTT 520
Db 513 ThrLeuAlaValAlaValAspAlaAsnValLeuIleAsnGluArgIleLysGlnGluLeu 532
Qy 521 TTATTGTCCTCAAGCTTAAATAATCTGTAGAAAAGATATACCAAGCTTTGGAGCC 580
Db 533 SerAsnGlyArgThrValGlnGlnAlaIleAspGluGlyTyrArgGlyAlaPheSerSer 552
Qy 581 ATTTTGTACTTACTGACTACGATATGGCTTCAGCACTTCTTCTTCTTACATACA 640
Db 553 IlePheAspAlaAsnIleThrThrLeuIleLysValIleIleLeuValAlaGlyThr 572
Qy 641 GGGCCTATTAAAGGTTGCTTGTGACATGATTAGGAATTTCTCTCAAGTTTACG 700
Db 573 GlyAlaIleLysGlyPheAlaIleThrThrGlyIleGlyValAlaThrSerMetPheThr 592
Qy 701 GCTCTTTCATGACTTAA 718
Db 593 AlaIleValGlyThrArg 598

RESULT 45
G85536
protein-export membrane protein secD [imported] - Escherichia coli (strain O157:H7, su
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 09-Nov-2001
C:Accession: G85536
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayh
iller, L.; Grobbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodac
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: G85536
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-615 <STO>
A:Cross-references: GB:AE005174; NID:q12513258; PIDN:AA954755.1; GSPDB:GN00145; UMGP:Z
A:Experimental source: strain O157:H7, substrain EDJ933
C:Genetics:
A:Gene: secD
C:Superfamily: protein export membrane protein secD

Alignment Scores:
Pred. No.: 3,23e-23 Length: 615
Score: 359.50 Matches: 80
Percent Similarity: 62.39% Conservative: 61
Best Local Similarity: 35.40% Mismatches: 82
Query Match: 10.85% Indels: 3
DB: 2 Gaps: 2

```

US-09-868-987-1 (1-1864) x G85536 (1-615)

```
Qy 44 AATGCACAATATCTGCACAAACCGTGGATGGCTGTAGTCATTGACGGTATATG 103
Db 375 SerGlyLysLysAspAlaAsn-----GlyAlaValLeuValLysGlnGluVal 392
Qy 104 GTCAGCAGCCCTATTTTAAACGTCCTCCATTGAAATCATGCCAGTGCTCAGGGAATTT 163
Db 393 IleAsnIleAlaAsnIleGlnSerArgLeuGlyAsnSerPheArgIleThrGlyIleAsn 412
Qy 164 ACCCACCCTGAAGTGAGCAAACTCGCCTCAGATTAAATCTGGAGCGATGCTCTTTTGT 223
Db 413 AsnProAsnGluAlaArgGlnLeuSerLeuLeuArgAlaGlyAlaLeuIleAlaPro 432
Qy 224 CCCGAGGTTCTCAGTGAGAGACCATCTCTTCATCTTGGGAAAAAACAATGTACACAA 283
Db 433 IleGlnIleValGluGluArgThrIleGlyProThrLeuGlyMetGlnAsnIleGluGln 452
Qy 284 GGCATTATCTCAGCATGCTGTGGCTTGGCAATGCTTATTTTGTAGAGCGTATATAT 343
Db 453 GlyLeuGluAlaCysLeuAlaGlyLeuValSerIleLeuPheMetIleIlePheTyr 472
Qy 344 AGATTGGAGCGCTCATCGCTTCGGAGCTGTCTCTCTGAATCTTTTCTTATCTGGSCA 403
Db 473 LysLysPheGlyLeuIleAlaThrSerAlaLeuIleAlaAsnLeuIleValGly 492
Qy 404 GCTCTCAGTATTG---GATGCGCCACTCACCTTGTTCAGGACTCGCTGGGATTGTTCTT 460
Db 493 IleMetSerLeuLeuProGlyAlaThrLeuSerMetProGlyIleAlaGlyIleValLeu 512
Qy 461 GCTATGGGATGGCGGTAGATGCAATGTTCTTGTATTTCGAAAGAAATCCGAGAGGAATTT 520
Db 513 ThrLeuAlaValAlaValAspAlaAsnValLeuIleAsnGluArgIleLysGluGluLeu 532
Qy 521 TTATTGTCTCAAGTCTTAAAAATCTGTAGAAAAAGGATATACCAGGCTTTTGGAGCC 580
Db 533 SerAsnGlyArgThrValGlnGlnAlaIleAspGluGlyTyrArgGlyAlaPheSerSer 552
Qy 581 ATTTTGTACTTAACCTGACTACATGATATTGGCCTCAGCACTTCTTTTCTCTAGATACA 640
Db 553 IlePheAspAlaAsnIleThrThrLeuIleLysValIleIleLeuTyrAlaValGlyThr 572
Qy 641 GGGCTATTAAAGGTTTGCTTGCATTTGATTTTGAATTTTCTCTCAATGTTTACG 700
Db 573 GlyAlaIleLysGlyPheAlaIleThrThrGlyIleGlyValAlaThrSerMetPheThr 592
Qy 701 GCTCTTTTTCATGACTAAA 718
Db 593 AlaIleValGlyThrArg 598
```

Search completed: December 17, 2002, 12:40:18
Job time : 91.5 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 17, 2002, 06:28:50 ; Search time 64 Seconds
(without alignments)
8931.961 Million cell updates/sec

Title: US-09-868-987-1
Perfect score: 1864
Sequence: 1 atggacttcgcattatgtc.....atgacaaattcagataatgc 1864

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA.*
1: /cgn2_6/ptodata/1/ina/5A.COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A.COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B.COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PCTUS.COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	116.4	6.2	6321	4	US-09-221-017B-311
2	99.6	5.3	2301	4	US-09-134-001C-852
3	84	4.5	2211	4	US-09-462-844-1
4	42.2	2.3	7218	1	US-08-232-463-14
5	39.6	2.1	1469	3	US-09-074-912-1
6	39.6	2.1	1469	4	US-09-230-136-1
C 7	38.8	2.1	298	4	US-09-071-035-147
C 8	38.8	2.1	441	4	US-09-071-035-145
9	37.2	2.0	3455	2	US-08-861-464-3
10	37.2	2.0	3455	2	US-08-396-001-3
11	37.2	2.0	3455	4	US-09-323-433A-3
12	37.2	2.0	8050	4	US-09-491-362-11
13	37.2	2.0	8050	4	US-09-874-562-11
14	36.2	1.9	499	4	US-09-641-638-52
15	36.2	1.9	499	4	US-09-641-638-572
16	35.8	1.9	246240	2	US-08-724-394A-20
17	35.8	1.9	246240	2	US-08-724-394A-21
18	35.8	1.9	246240	2	US-08-724-394A-22
19	35.4	1.9	19124	2	US-08-487-826B-13
C 20	34.8	1.9	1245	4	US-09-134-001C-2535
21	34.8	1.9	5529	3	US-08-869-696-1
22	34.4	1.8	13121	4	US-08-961-527-126
23	33.8	1.8	38844	4	US-09-734-675-3
24	33.4	1.8	3893	4	US-09-221-017B-604
C 25	33.2	1.8	1023	2	US-08-757-653-175
C 26	33.2	1.8	1023	2	US-08-823-516-78
C 27	33.2	1.8	1023	3	US-08-759-038-114

C	28	33.2	1.8	1023	3	US-08-758-314-114	Sequence 114, App
	29	33.2	1.8	2476	4	US-08-844-274-11	Sequence 11, Appl
	30	33.2	1.8	5194	4	US-08-844-274-16	Sequence 16, Appl
	31	33.2	1.8	5194	4	US-08-844-274-17	Sequence 17, Appl
C	32	33.2	1.8	5679	4	US-08-844-274-10	Sequence 10, Appl
	33	33.2	1.8	6448	4	US-08-844-274-15	Sequence 15, Appl
C	34	33.2	1.8	6723	4	US-08-844-274-13	Sequence 13, Appl
	35	33.2	1.8	6723	4	US-08-844-274-14	Sequence 14, Appl
	36	33.2	1.8	7560	4	US-08-844-274-20	Sequence 20, Appl
C	37	33.2	1.8	8898	4	US-08-961-527-69	Sequence 69, Appl
	38	33	1.8	2304	1	US-08-464-266-1	Sequence 1, Appl
C	39	33	1.8	2304	1	US-08-464-272-1	Sequence 1, Appl
C	40	33	1.8	2304	4	US-08-464-514-1	Sequence 1, Appl
	41	33	1.8	2304	4	US-08-486-403-1	Sequence 1, Appl
C	42	32.6	1.7	504	4	US-09-149-476-221	Sequence 221, App
	43	32.4	1.7	1260	3	US-09-008-979A-3	Sequence 3, Appl
	44	32.4	1.7	1260	4	US-09-460-618-3	Sequence 3, Appl
	45	32.4	1.7	1260	4	US-09-310-235B-3	Sequence 3, Appl

ALIGNMENTS

RESULT 1
US-09-221-017B-311/c
; Sequence 311, Application US/09221017B
; Patent No. 6444799
; GENERAL INFORMATION:
; APPLICANT: ROSS, Bruce C.
; TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF
; NUMBER OF SEQUENCES: 1120
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 755 PAGE MILL ROAD
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304-1018
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows
; SOFTWARE: FastSeq for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/221,017B
; FILING DATE: 23-DEC-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PP1182
; FILING DATE: 31-DEC-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PP1546
; FILING DATE: 30-JAN-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PP2911
; FILING DATE: 09-APR-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/AU98/01023
; FILING DATE: 10-DEC-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Monroy, Gladys H
; REGISTRATION NUMBER: 32,430
; REFERENCE/DOCKET NUMBER: 27340-20021.00
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-813-5600
; TELEFAX: 650-494-0792
; TELEX: 706141
; INFORMATION FOR SEQ ID NO: 311:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6321 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular

MOLECULE TYPE: DNA (genomic)
 ;
 ; HYPOTHEICAL: NO
 ; ANTI-SENSE: UNKNOWN
 ; ORIGINAL SOURCE:
 ; ORGANISM: PORYPHYROMONAS GINGIVALIS
 ; FEATURE:
 ; NAME/KEY: misc feature
 ; LOCATION: 1...6321
 ; US-09-221-017B-311

Query Match 6.2%; Score 116.4; DB 4; Length 6321;
 Best Local Similarity 50.8%; Pred. No. 7.7e-22;
 Matches 304; Conservative 0; Mismatches 231; Indels 3; Gaps 1;

QY 150 TCTCAGGAAATTTACCAACCGTGAAGTGAAGAACTGCGCTCAGATTAAATCTGAG 209
 DB 2009 TCTCCGGGACCTTCAACCTGAGAGAGCGCGGTGACCTTGCCACAGTACTCCGATA 1950
 QY 210 CGATGCTTTTGTCCGAGGTTCTCAGTGAAGAGACATCTTTGATCTTGGGAAA 269
 DB 1949 AAATGAGATGCTACGATGAAGATCGAACAAGAAAGTATGTTCTTACGCTGGGTCGG 1890
 QY 270 AACATGACACAAGGCAATTAATCAGCATGCTGAGCTTGGCAATGCTTATTTGTTGA 329
 DB 1889 AGTCATTAAAGCCGATTTCTGTCTCCGCTCGCTTGGTTATCCTGATGTGTACA 1830
 QY 330 TGAACGATATTA--TAGATTTGAGAGGCTCATGCTTCGGAGCTGTTCTTGAATC 386
 DB 1829 TGTGTGCTTACGAGTTTCTTGCCGGGCTTATCGCAAAAGCGGCATGATGTGTAACA 1770
 QY 387 TTTTGCTATCTGGGAGCTCTACAGTATTGGATGCGCCACTACCTTTCAGAGACTCG 446
 DB 1769 GCTTCTTACATGAGGCGTATTTGCTTTTTCATGCGCTGACCTCTCGGGTATCG 1710
 QY 447 CTGGGATTTGTTCTTCATGAGGGAATGCGCGTAGATGCAATGTTCTTATTCGAAAGA 506
 DB 1709 CAGGTTTGCTCTACGCTGGGTATGCTGTGAGATGCCAAGTACTTATCTTCAGCGTA 1650
 QY 507 TCCGAGAGAAATTTTATTTGCTCAAGTCTTAAAAATCTGTAAAAAGGATATACCA 566
 DB 1649 TCAAGAAGAGCTTCGTCCGGGTAAAGACTCGAATTCGCGCTTACGAGATGTTATGCA 1590
 QY 567 AGGCTTTTGGAGCAATTTTATCTTACTGACTACAGTATGAGCCCTCAGACACTCTCT 626
 DB 1589 ACGCTTCTCTCGCATCTTCACTCGAAGCTTACGACTATATTAACGGATATCACTCAT 1530
 QY 627 TCTTCTAGATACAGGCGCTATTAAAGGTTGCTTTGACATGATTTTGAATTTTCT 686
 DB 1529 TCCCTACAGGAGCGGGCGCATTCGCGGTTTGCACATCAGTTGATTAATCGGTCTTATCG 1470
 QY 687 CTTCATGTTTACGCGCTCTTTTCACTGACTTAAATTTTCTTCATGCTGTGATGAATA 744
 DB 1469 CTTCTTCAATACGCGCTCTCTTCTTGACTGATGCTTTCGAGAAACTCGCGCAAAA 1412

RESULT 2
 US-09-134-001C-852
 ; Sequence 852, Application US/09134001C
 ; Patent No. 6380370
 ; GENERAL INFORMATION:
 ; APPLICANT: Lynn Doucette-Stamm et al
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
 ; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: GTC-007
 ; CURRENT APPLICATION NUMBER: US/09/134,001C
 ; CURRENT FILING DATE: 1998-08-13
 ; PRIOR APPLICATION NUMBER: US 60/064,964
 ; PRIOR FILING DATE: 1997-11-08
 ; PRIOR APPLICATION NUMBER: US 60/055,779
 ; PRIOR FILING DATE: 1997-08-14
 ; NUMBER OF SEQ ID NOS: 5674
 ; SEQ ID NO 852
 ; LENGTH: 2301

TYPE: DNA
 ; ORGANISM: Staphylococcus epidermidis
 ; US-09-134-001C-852

Query Match 5.3%; Score 99.6; DB 4; Length 2301;
 Best Local Similarity 48.7%; Pred. No. 1.8e-17;
 Matches 270; Conservative 0; Mismatches 284; Indels 0; Gaps 0;

QY 173 GAAGTGAACAACCTCGCTCAGATTTAATAATCTGAGCCAGTGTCTTTGTTCCCGAGGTT 232
 DB 694 GAAGCGAACAATAATGATGATTAAATGCGCGCTCATACAGTGTGATTTAAAGAA 753
 QY 233 CTGAGTAAGAGAGCATCTCTTGATCTTGGGAAAAACAATGTACAGAGCATTAAC 232
 DB 754 ATTTACTTACCTCTGTGGTGCACATTTGGTCAAGATGCTTTGATTAAGACATGTTT 813
 QY 293 TCAAGATGCTGTGGCTTGGCAATGCTTAATTTGATGAGCTATATTAATTAATTTGGA 352
 DB 814 GCATCATTTGTAGTATAGCATTAATTTATTTATGCTTGTTCATCGTTTGCTT 873
 QY 353 GCGGTATCGCTTGGGAGCTGTTCTTGAAATCTTTGCTTATCTGGGAGCTCTACAG 412
 DB 874 GGTATGATTGCAATCAATGCTTACCACTTATATTTAATTTAATTAAGTGCATTCAT 933
 QY 413 TATTTGATGCGCACCTGACCTGTGAGACTGCGTGGGATTTGTTCTGTATGGGATG 472
 DB 934 TTCAATACAGGTATTAATCTTACCTGATGCGGCAATTAATTTAGGTGATGATG 933
 QY 473 GCGGTAGATCAATATGTTCTTGTATTCGAAAGATCCGAGAGAAATTTTATGTTCTCA 532
 DB 994 GCTGTGATCCCAATATCAATATGATGAAAGATTTAAAGATGAACTTAAGAAATTTGA 1053
 QY 533 AGTCTTAAAAATCTGTGAAAAAGATATACCAAGGCTTTTGGAGCCATTTTGAATCT 552
 DB 1054 ACCGTTAAACAACGATTAACAAAGCAATAAAGTTCAATTTCAATTAATTTGATTC 1113
 QY 593 AACTGATCTACATTAATGGCTCAGACCTTTTCTTCTAATACAGGCGCTATTA 652
 DB 1114 AACTTAACAATCTCATCGCTGAGCTGTGCTTTCTTCTTGAAGAAAGTTAGTCAAA 1173
 QY 653 GGGTTGCTTTGACATTTAGTAATTTCTCTTCAATGTTTACGCTCTTTTCATG 712
 DB 1174 GGCTTGCAACCAATGTTACTTAGGATTTTATGATATTTGTAACCGAGATTCCTTA 1233
 QY 713 ACTTAATTTTCTT 726
 DB 1234 TCAAGAGGGTGT 1247

RESULT 3
 US-09-462-844-1
 ; Sequence 1, Application US/09462844
 ; Patent No. 6258563
 ; GENERAL INFORMATION:
 ; APPLICANT: Quax, Wilhelmus J.
 ; TITLE OF INVENTION: Increasing Production of Proteins in
 ; TITLE OF INVENTION: Gram-Positive Microorganisms
 ; FILE REFERENCE: GC385-US
 ; CURRENT APPLICATION NUMBER: US/09/462,844
 ; CURRENT FILING DATE: 2000-01-13
 ; PRIOR APPLICATION NUMBER: PCT/US98/14786
 ; PRIOR FILING DATE: 1998-07-15
 ; PRIOR APPLICATION NUMBER: EP 97305286.3
 ; PRIOR FILING DATE: 1997-07-16
 ; PRIOR APPLICATION NUMBER: EP 97305344.0
 ; PRIOR FILING DATE: 1997-07-17
 ; NUMBER OF SEQ ID NOS: 11
 ; SOFTWARE: FaSTSeq for Windows Version 4.0
 ; SEQ ID NO 1
 ; LENGTH: 2211
 ; TYPE: DNA
 ; ORGANISM: Bacillus subtilis
 ; US-09-462-844-1

Query Match 4.5%; Score 84; DB 4; Length 2211;
Best Local Similarity 49.8%; Pred. No. 3e-13;
Matches 213; Conservative 0; Mismatches 215; Indels 0; Gaps 0;

Qy 305 GCGTGGCAATGCTTATTTGATGAGCGTATATATAGATTGGAGGCGTCACTGCT 364
Db 793 GGTATCGCAATATTTCTTATTTATGCTTTCTATACCGTCTCCGGGATTAATCGG 852
Qy 365 TCGGAGCTGTCTTCTGAATCTTTGCTTATCTGGGCGAGCTCTACAGTATTTGGATCG 424
Db 853 GTGATTACGCTGTCTGTTATATCTACATTACACTCCAGATCTTTGACTGGATGAATGCC 912
Qy 425 CCACTACCTTGTGAGGACTGCTGGGATTTCTTGTATGCGGATGCGCGTAGATGCA 484
Db 913 GTACTACGCTTCGGGAATTCGCGCTCTCATTTTATGCTGGGATGCGTGTGACGCC 972
Qy 485 AATGTTCTTGTATTCGAAGAAATCGAGAGAAATTTTATGCTCAAGTCTTAAAAA 544
Db 973 AACATTATCTATGAGCGGATTAAGAAGAGCTCAAGTAGGAAATCAGTCGCTCT 1032
Qy 545 TCTGAGAAAAGGATATACCAAGCTTTTGGAGCCATTTTGTATTTCAACTTCACTACA 604
Db 1033 GCCTTCGCTTCAGAAACAGAGGTCATTTGCGAGATTTTGGCGGAATATTACACC 1092
Qy 605 GTATGGCTCAGCACTCTTTCTCTCTAGATACAGGCGCTATTAAGGTTGCTTTG 664
Db 1093 ATTATGCGGGTGTGCTCTTTATCTTTGGGCAAGCTCTGTTAAAGGTTTGGACA 1152
Qy 665 ACATTGATTTAGGAATTTCTCTCATGTTTACGCTCTTTTCATGATTAATTTTC 724
Db 1153 ATGCTGATCTATCGATTTTGAAGCTTTATCAAGCTTTTCTTATCGAGATTTCTC 1212
Qy 725 TTCATGCT 732
Db 1213 CTCGCTCT 1220

RESULT 4
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOMLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232.463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935.313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU

TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-F18
; US-08-232-463-14

Query Match 2.3%; Score 42.2; DB 1; Length 7218;
Best Local Similarity 8.6%; Pred. No. 0.11;
Matches 38; Conservative 206; Mismatches 199; Indels 0; Gaps 0;

Qy 1405 GAAGAAATTCAAATAGATTTCGAGCATTTGGTCTTTTAATGACTGTATTTGGGTATTC 1464
Db 1024 GAATTAATTCGAGCTTGGCTGCGAGTTCGAGGAGCTTGCATTTTTTTTTTTTTT 1083
Qy 1465 ATTAACAATACCTTTGATCATTTTGTATCGTATTCGTAAGATCGCAAGCGAACCTGTT 1524
Db 1084 YY 1143
Qy 1525 TACCCCTATGATGTTTATGTTATGATGCTTCAAAAGACGTTTCAGCCGACGGTAAT 1584
Db 1144 YY 1203
Qy 1585 GACAAGACTACACTATCAGCTTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1644
Db 1204 YY 1263
Qy 1645 CTTTAATTTGCAATTTATGACCATAGGATTTCTTAGGAACCTTTATGCTCTCTTA 1704
Db 1264 YY 1323
Qy 1705 TATTCACCACTCTGTTGTTTATGTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1764
Db 1324 YY 1383
Qy 1765 GTTAACTTAACTTAAGTGTAGCAATATAAAATCTCCTTTTGGAGCTTTAGTCCCAAG 1824
Db 1384 YY 1443
Qy 1825 GCCCTGCTGTTAATTTATG 1847
Db 1444 TCTTCTATCTCTTTAACTACTTG 1466

RESULT 5
US-09-074-912-1
; Sequence 1, Application US/09074912
; Patent No. 6057491
; GENERAL INFORMATION:
; APPLICANT: CIGAN, AMY L
; APPLICANT: CZAPLA, THOMAS H
; APPLICANT: FALLIS, LYNN
; APPLICANT: MEYER, TERRY E
; APPLICANT: MUNDELL, SCOTT A
; APPLICANT: SABUS, BRIAN
; APPLICANT: SCHUBERT, KAREL
; TITLE OF INVENTION: PROTEINS HAVING INSECTICIDAL ACTIVITIES
; TITLE OF INVENTION: AND METHODS OF USE
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: W. MURRAY SPRULL (ALSTON & BIRD, LLP)
; STREET: 3605 GLENWOOD AVE.
; CITY: RALEIGH
; STATE: NC
; COUNTRY: USA
; ZIP: 27622

```
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/074,912
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: SPRULL, W. MURRAY
REGISTRATION NUMBER: 32,943
REFERENCE/DOCKET NUMBER: 5718-9
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919 420 2202
TELEFAX: 919 881 3175
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1469 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
ORIGINAL SOURCE:
ORGANISM: Pentaclethra macroloba
FEATURE:
NAME/KEY: CDS
LOCATION: 31..1257
US-09-074-912-1

Query Match      2.1%; Score 39.6; DB 3; Length 1469;
Best Local Similarity 50.5%; Pred. No. 0.28; Mismatches 94; Indels 0; Gaps 0;
Matches 96; Conservative 0;

QY 1315 TGAATGCAATATGCTTTCAGTGCCTTATGCGCTTATTCATGACCTTTGGCTACCTG 1374
DB 678 TGAATTCATCTCGTTGATGTCGATCATGCTGATTTCCGCCCGGTTGCTTCAG 737
QY 1375 TGCAGTCTTGTATTAGACATTTCTTTGAAGAAATTAATTAATTTGCAAGCCAT 1434
DB 738 CGAGTGCTCCAGCAAGAAATAACAAGATAAGAAATCTTTGCTGCTATAGGAAC 797
QY 1435 TGTGCTTTATGACTGATTTGGGCTATTCTTAACAATGCTTGAATCATTTTGAATG 1494
DB 798 TGGAGTTGTAAACCTGCTGAGGGTTATTCCTATGCTGACTGACTATTTTCGATTG 857
QY 1495 TATTCGTGAA 1504
DB 858 GAGTAGTGAA 867

RESULT 6
US-09-290-136-1
Sequence 1, Application US/09290136
Patent No. 6339144
GENERAL INFORMATION:
APPLICANT: Cigan, Amy L.
APPLICANT: Czapl, Thomas H.
APPLICANT: Fallis, Lynn
APPLICANT: Meyer, Terry E.
APPLICANT: Mundell, Scott A.
APPLICANT: Sabus, Brian
APPLICANT: Schubert, Karen
TITLE OF INVENTION: Proteins Having Insecticidal Activities and Methods of
TITLE OF INVENTION: Use
FILE REFERENCE: 5718-9A, 035718/180486
CURRENT APPLICATION NUMBER: US/09/290,136
CURRENT FILING DATE: 1999-04-13
EARLIER APPLICATION NUMBER: 09/074,912
EARLIER FILING DATE: 1998-05-08
NUMBER OF SEQ ID NOS: 11
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 1
```

```
LENGTH: 1469
TYPE: DNA
ORGANISM: Pentaclethra macroloba
FEATURE:
NAME/KEY: CDS
LOCATION: (31)..(1257)
US-09-290-136-1

Query Match      2.1%; Score 39.6; DB 4; Length 1469;
Best Local Similarity 50.5%; Pred. No. 0.28; Mismatches 94; Indels 0; Gaps 0;
Matches 96; Conservative 0;

QY 1315 TGAATGCAATATGCTTTCAGTGCCTTATGCGCTTATTCATGACCTTTGGCTACCTG 1374
DB 678 TGAATTCATCTCGTTGATGTCGATCATGCTGATTTCCGCCCGGTTGCTTCAG 737
QY 1375 TGCAGTCTTGTATTAGACATTTCTTTGAAGAAATTAATTAATTTGCAAGCCAT 1434
DB 738 CGAGTGCTCCAGCAAGAAATAACAAGATAAGAAATCTTTGCTGCTATAGGAAC 797
QY 1435 TGTGCTTTATGACTGATTTGGGCTATTCTTAACAATGCTTGAATCATTTTGAATG 1494
DB 798 TGGAGTTGTAAACCTGCTGAGGGTTATTCCTATGCTGACTGACTATTTTCGATTG 857
QY 1495 TATTCGTGAA 1504
DB 858 GAGTAGTGAA 867

RESULT 7
US-09-071-035-147/c
Sequence 147, Application US/09071035
Patent No. 6448043
GENERAL INFORMATION:
APPLICANT: Gil H. Choi
TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides
NUMBER OF SEQUENCES: 496
CORRESPONDENCE ADDRESS:
ADDRESSER: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.44M storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/071,035
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: A. Anders Brooks
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB369P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 147:
SEQUENCE CHARACTERISTICS:
LENGTH: 298 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-09-071-035-147

Query Match      2.1%; Score 38.8; DB 4; Length 298;
Best Local Similarity 50.0%; Pred. No. 0.23;
```



```

QY 1713 CACCTCTGTTGTTGTTATGTCGCTAAAGAAAAATCGCTCAAAAATAAGTACCGTTAAACT 1772
DB 277 TATCACTAAATTTTGAATTATACCTCGAAAAAGTTTCATCATCAAAAAAAGTGTAAAGTAG 218
QY 1773 TAATCTAAGCTGTA 1786
DB 217 TAGTTGGACTTTTA 204

RESULT 9
US-08-861-464-3
; Sequence 3, Application US/08861464
; Patent No. 5874210
; GENERAL INFORMATION:
; APPLICANT: Guarente, Leonard P.
; APPLICANT: Austriaco Jr., Nicanor
; APPLICANT: Kennedy, Brian
; TITLE OF INVENTION: Genes Determining Cellular Senescence
; TITLE OF INVENTION: in Yeast
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: MA
; COUNTRY: USA
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/861,464
; FILING DATE: 22-MAY-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/396,001
; FILING DATE: 28-FEB-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/09351
; FILING DATE: 15-AUG-1994
; APPLICATION DATA:
; APPLICATION NUMBER: US 08/107,408
; FILING DATE: 16-AUG-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: MIT-6408A22
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 781-861-6240
; TELEFAX: 781-861-9540
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3455 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 663..3164
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 663..3164
; US-08-861-464-3

Query Match 2.0%; Score 37.2; DB 2; Length 3455;
Best Local Similarity 57.9%; Pred. No.1.8;
Matches 66; Conservative 0; Mismatches 48; Indels 0; Gaps 0
QY 522 TATTGTCTCAAAAGTCTTAAAAAATCTGTAGAAAAAGGATATACCAAGCGCTTTTGGAGCCA 581

```

Db 2207 TAGTACGCAATTGGGCAAAACAGCTCTAGTACATTAAGGTAACCTTTGTAACGA 2266
Qy 582 TTTTGAATTCATCTGACTAGACATATTGGGCTCAGCACTTTCTTCTCTAG 635
Db 2267 TTTTTCATTTGAAATTTGGTAACCTTGTCTTCAAAAGAAATTAATCTTCCAG 2320

RESULT 10
US-08-396-001-3
Sequence 3, Application US/08396001
Patent No. 5919618
GENERAL INFORMATION:
APPLICANT: Guarente, Leonard P.
APPLICANT: Austriaco Jr., Nicanor
APPLICANT: Claus, James
APPLICANT: Cole, Francesca
APPLICANT: Kennedy, Brian
TITLE OF INVENTION: Genes Determining Cellular Senescence in
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
STREET: Two Militia Drive
CITY: Lexington
STATE: MA
COUNTRY: USA
ZIP: 02173
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/396,001
FILING DATE: 28-FEB-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Granahan, Patricia
REGISTRATION NUMBER: 32,227
REFERENCE/DOCKET NUMBER: MIT-6408A2
TELEPHONE: 617-861-6240
TELEFAX: 617-861-9540
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 3455 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 663..3164
US-08-396-001-3

Query Match 2.0%; Score 37.2; DB 2; Length 3455;
Best Local Similarity 57.9%; Pred. No. 1.8;
Matches 66; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

Qy 522 TATGTCTCAAGCTTTAAATAATCTGAGAAAAGATATACCAAGCTTTGGAGCCA 581
Db 2207 TAGTACGCAATTGGGCAAAACAGCTCTAGTACATTAAGGTAACCTTTGTAACGA 2266

Qy 582 TTTTGAATTCATCTGACTAGACATATTGGGCTCAGCACTTTCTTCTCTAG 635
Db 2267 TTTTTCATTTGAAATTTGGTAACCTTGTCTTCAAAAGAAATTAATCTTCCAG 2320

RESULT 11
US-09-323-433A-3
Sequence 3, Application US/09323433A
Patent No. 6218512
GENERAL INFORMATION:

APPLICANT: Guarente, Leonard P.
APPLICANT: Austriaco Jr., Nicanor
APPLICANT: Claus, James J.
APPLICANT: Cole, Francesca
APPLICANT: Kennedy, Brian
TITLE OF INVENTION: GENES DETERMINING CELLULAR SENESCENCE IN
NUMBER OF SEQUENCES: 16
FILE REFERENCE: 0050,1491-003
CURRENT APPLICATION NUMBER: US/09/323,433A
CURRENT FILING DATE: 1999-06-01
PRIOR APPLICATION NUMBER: US 08/396,001
PRIOR FILING DATE: 1995-02-28
PRIOR APPLICATION NUMBER: PCT/US94/09351
PRIOR FILING DATE: 1994-08-15
PRIOR APPLICATION NUMBER: US 08/107,408
PRIOR FILING DATE: 1993-08-16
NUMBER OF SEQ ID NOS: 48
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 3
LENGTH: 3455
TYPE: DNA
ORGANISM: Saccharomyces cerevisiae
FEATURE:
NAME/KEY: CDS
LOCATION: (663)...(3164)
OTHER INFORMATION: UTH4
US-09-323-433A-3

Query Match 2.0%; Score 37.2; DB 4; Length 3455;
Best Local Similarity 57.9%; Pred. No. 1.8;
Matches 66; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

Qy 522 TATGTCTCAAGCTTTAAATAATCTGAGAAAAGATATACCAAGCTTTGGAGCCA 581
Db 2207 TAGTACGCAATTGGGCAAAACAGCTCTAGTACATTAAGGTAACCTTTGTAACGA 2266

Qy 582 TTTTGAATTCATCTGACTAGACATATTGGGCTCAGCACTTTCTTCTCTAG 635
Db 2267 TTTTTCATTTGAAATTTGGTAACCTTGTCTTCAAAAGAAATTAATCTTCCAG 2320

RESULT 12
US-09-491-362-11
Sequence 11, Application US/09491362
Patent No. 6281017
GENERAL INFORMATION:
APPLICANT: Croteau, Rodney B
APPLICANT: Lange, Bernd M
TITLE OF INVENTION: 1-DEOXY-D-XYLULOSE-5-PHOSPHATE REDUCTOISOMERASE, AND
TITLE OF INVENTION: METHODS OF USE
FILE REFERENCE: WSUR14977
CURRENT APPLICATION NUMBER: US/09/491,362
CURRENT FILING DATE: 2000-01-26
EARLIER APPLICATION NUMBER: 60/118,349
EARLIER FILING DATE: 1999-02-03
NUMBER OF SEQ ID NOS: 13
SOFTWARE: Patent In Ver. 2.0
SEQ ID NO 11
LENGTH: 8050
TYPE: DNA
ORGANISM: Arabidopsis thaliana
US-09-491-362-11

Query Match 2.0%; Score 37.2; DB 4; Length 8050;
Best Local Similarity 53.4%; Pred. No. 2.7;
Matches 78; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

Qy 1363 TTGGCTACCTGGGAGCTGTGTTATAGACATTTCTTTGAGAAAATTCAAATAGA 1422
Db 7423 TTGGTCCCGGTGGGTTTGTACATAGTAGATTTTATATAGAGCTGTAAATGAGT 7482

Qy 1423 TTGCAAGCACTGGTGCTTTAATGACTGATTTGGGGTATTCATTAACAATCTTGAT 1482

Db	7483	TCAGCAAGTTTACCATTTGCTCAGGAAACTGTTATGAGGATCCTCCAAACGTCTCTGTTTAA	7542
Qy	1483	CATTTTTCGATTCGTATTCGTGAAGATC	1508
Db	7543	GAATTCAGTACCAATTCGAGAGGATC	7568

```

RESULT 13
US-09-874-562-11
; Sequence 11, Application US/09874562
; Patent No. 6420159
; GENERAL INFORMATION:
; APPLICANT: Croteau, Rodney B
; APPLICANT: Lange, Bernd M
; TITLE OF INVENTION: 1-DEOXY-D-XYLULOSE-5-PHOSPHATE REDUCTOISOMERASE, AND
; TITLE OF INVENTION: METHODS OF USE
; FILE REFERENCE: WSUR17549
; CURRENT APPLICATION NUMBER: US/09/874,562
; CURRENT FILING DATE: 2001-06-04
; PRIOR APPLICATION NUMBER: 09/491,362
; PRIOR FILING DATE: 2000-01-26
; PRIOR APPLICATION NUMBER: 60/118,349
; PRIOR FILING DATE: 1999-02-03
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
; LENGTH: 8050
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-874-562-11

```

	Query Match	2.0%;	Score 37.2;	DB 4;	Length 8050;
	Best Local Similarity	53.4%;	Pred. No. 2.7;		
	Matches 78;	Conservative 0;	Mismatches 68;	Indels 0;	Gaps 0;
QY	1363	TTTGGCTACCTGTCGACGCTTGTTTATAGCACATTTCTTTTGAAGAAAATCAAATAGA	1422		
Db	7423	TTTGGTTCCCGGTTTCGGTGTGTCATAGTAGGAGATTTTATAAAGCCCTGCTAATGAGGT	7482		
QY	1423	TTTGAAGGCCATTCGGTGGTGTAAATGACTGTATTCGGGGTATTCATTAAACAATACCTTTCAT	1482		
Db	7483	TCAGCAAGTTACCATTCGCTCAGGAAACGTGTATGGAGATCCTCCAAGCTCTCTGTTTAA	7542		
QY	1483	CATTTTTGATCGTATTCGTGAAGATC	1508		
Db	7543	GAATTCAGTACCAATTCGAGAGGATC	7568		

RESULT 14
 US-09-641-638-52
 ; Sequence 52, Application US/09641638
 ; Patent No. 6432648
 ; GENERAL INFORMATION:
 ; APPLICANT: Blumenfeld, Marta
 ; APPLICANT: Bougueleret, Lydie
 ; APPLICANT: Chumakov, Ilya
 ; APPLICANT: Cohen, Annick
 ; TITLE OF INVENTION: BIALLELIC MARKERS DERIVED FROM GENOMIC REGIONS CARRYING
 ; TITLE OF INVENTION: GENES INVOLVED IN ARACHIDONIC ACID METABOLISM
 ; FILE REFERENCE: GENSET.051CPI
 ; CURRENT APPLICATION NUMBER: US/09/641.638
 ; CURRENT FILING DATE: 2000-08-16
 ; PRIOR APPLICATION NUMBER: US 09/502,330
 ; PRIOR FILING DATE: 2000-02-11
 ; PRIOR APPLICATION NUMBER: US 60/133,200
 ; PRIOR FILING DATE: 1999-05-07
 ; PRIOR APPLICATION NUMBER: US 09/275,267
 ; PRIOR FILING DATE: 1999-03-23
 ; PRIOR APPLICATION NUMBER: US 60/119,917
 ; PRIOR FILING DATE: 1999-02-12
 ; NUMBER OF SEQ ID NOS: 1304
 ; SOFTWARE: Patent.pm
 ; SEQ ID NO 52

```

; LENGTH: 499
; TYPE: DNA
; ORGANISM: Homo Sapiens
; FEATURE:
; NAME/KEY: allele
; LOCATION: 466
; OTHER INFORMATION: 12-215-467 : polymorphic base G or T
; NAME/KEY: misc_binding
; LOCATION: 446..465
; OTHER INFORMATION: 12-215-467.mis1, potential
; NAME/KEY: misc_binding
; LOCATION: 467..486
; OTHER INFORMATION: 12-215-467.mis2, potential complement
; NAME/KEY: primer_bind
; LOCATION: 1..20
; OTHER INFORMATION: upstream amplification primer
; NAME/KEY: primer_bind
; LOCATION: 479..499
; OTHER INFORMATION: downstream amplification primer, complement
; NAME/KEY: misc_binding
; LOCATION: 454..478
; OTHER INFORMATION: 12-215-467 potential probe
; NAME/KEY: misc_feature
; LOCATION: 87
; OTHER INFORMATION: n=a, g, c or t
; US-09-641-638-52

Query Match          1.9%; Score 36.2; DB 4; Length 499;
Best Local Similarity 50.2%; Pred. No. 1.5;
Matches 115; Conservative 0; Mismatches 113; Indels 1; Gaps 1;

Qy 1322 CAATATGCTTTTCAGTGCCTATGCGCTTTAATTCATGACCTTTGGCTACCTGTGCAGTC 1381
Db 217 CATTTTGAGTTGACTTTTGTGTATGTTATAAATAAGCTCTAAATTTTCATTTTTTTTAAATGT 276
Qy 1382 TTGTTTATAGCACATTTCTTTTTTGAAGAAAAATTCAAATAGAT-TTGCAGCCATTGGTGC 1440
Db 277 TTATGTTTCAGCTCTCTTTTGTGTTTGTGCTCACAAATTTCTTTAGATATTAGAAGTC1TTTGTAG 336
Qy 1441 TTTAATGACTGTATTGGGGTGATTTCATTAAACAATAC1TTTGATCATTTTTCATGCTGATTTCG 1500
Db 337 CTCATAACAAGTTTATGAGATTTCTTTTTTCTATTTCAGTGAAAAATGCCATTGGAATTTT 396
Qy 1501 TGAAGATCGCAACGCAACCTGTTTACCCCTATGCAATGTTTTAGTTAAT 1549
Db 397 GAAAGGGGTTTGACCGCAATCTTTAGATACATTTTCGGGTAGTAGACACATT 445

```

```

RESULT 15
US-09-641-638-572
; Sequence 572, Application US/09641638
; Patent No. 6432648
; GENERAL INFORMATION:
; APPLICANT: Blumenfeld, Marta
; APPLICANT: Bougueleret, Lydie
; APPLICANT: Chumakov, Ilya
; APPLICANT: Cohen, Annick
; TITLE OF INVENTION: BIALLELIC MARKERS DERIVED FROM GENOMIC REGIONS CARRYING
; TITLE OF INVENTION: GENES INVOLVED IN ARACHIDONIC ACID METABOLISM
; FILE REFERENCE: GENSET.051CPI
; CURRENT APPLICATION NUMBER: US/09/641,638
; CURRENT FILING DATE: 2000-08-16
; PRIOR APPLICATION NUMBER: US 09/502,330
; PRIOR FILING DATE: 2000-02-11
; PRIOR APPLICATION NUMBER: US 60/133,200
; PRIOR FILING DATE: 1999-05-07
; PRIOR APPLICATION NUMBER: US 09/275,267
; PRIOR FILING DATE: 1999-03-23
; PRIOR APPLICATION NUMBER: US 60/119,917
; PRIOR FILING DATE: 1999-02-12
; NUMBER OF SEQ ID NOS: 1304
; SOFTWARE: patent.pm
; SEQ ID NO 572

```

LENGTH: 499
TYPE: DNA
ORGANISM: Homo Sapiens
FEATURE:
NAME/KEY: allele
LOCATION: 271
OTHER INFORMATION: 12-215-272 : deletion T
NAME/KEY: misc_binding
LOCATION: 251..270
OTHER INFORMATION: 12-215-272.misl, potential
NAME/KEY: primer_bind
LOCATION: 1..20
OTHER INFORMATION: upstream amplification primer
NAME/KEY: primer_bind
LOCATION: 479..499
OTHER INFORMATION: downstream amplification primer, complement
NAME/KEY: misc_feature
LOCATION: 87
OTHER INFORMATION: n=a, g, c or t
US-09-641-638-572

Query Match 1.9%; Score 36.2; DB 4; Length 499;
Best Local Similarity 50.2%; Pred. No. 1.5;
Matches 115; Conservative 0; Mismatches 113; Indels 1; Gaps 1;

QY 1322 CAATATGCTTCAAGTCCGATGCGCTTTAATTCATGACCTTTGGCTACCTGCGACGTC 1381
DB 217 CATTTGGAGTTGACTTTTGTGATGTTAAATAGGCTCTAATTTCTTTTAAATGT 276

QY 1382 TTGTTATAGACATTTCTTTTGAAGAAATTCAAATAGAT-TTGAAGCATGTGTC 1440
DB 277 TTAAGTCAGCTCTTTGTTTGTCTCAATTTCTTTAGATTAAGAGCTTTTGTAG 336

QY 1441 TTTAATGACTGATGGGATATTCATTAACATATCTTGATTCATTTTGAATCGTATTCG 1500
DB 337 CTCATATACAGTTTAGATTTCTTTTCTATTTTCAGTGAAGAAATGCAATTTGAAATTTT 396

QY 1501 TGAAGATCCCAAGCAACCTGTTTACCCCTATGATGATTTTATTTAT 1549
DB 397 GAAAGGGGTCAGCGAATCTTTAGATCATTGGGTAGTAGACATTT 445

RESULT 16
US-08-724-394A-20
Sequence 20, Application US/08724394A
Patent No. 5872237
GENERAL INFORMATION:
APPLICANT: Feder, John N.
APPLICANT: Krommal, Gregory S.
APPLICANT: Laufer, Peter M.
APPLICANT: Ruddy, David A.
APPLICANT: Thomas, Winston
APPLICANT: Tsuchihashi, Zenta
APPLICANT: Wolff, Roger K.
TITLE OF INVENTION: Megabase Transcript Map: No. 5872237el
TITLE OF INVENTION: Sequences and Antibodies Thereo
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP
STREET: Two Embarcadero Center, 8th Floor
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/724,394A
FILING DATE: 01-OCT-1996
CLASSIFICATION: 536

ATTORNEY/AGENT INFORMATION:
NAME: Filts, Renee A.
REGISTRATION NUMBER: 35,136
REFERENCE/DOCKET NUMBER: 017957-000100
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-576-0200
TELEFAX: 415-576-0300
INFORMATION FOR SEQ. ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 246240 base pairs
TYPE: nucleic acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1..246240
OTHER INFORMATION: /note= "HLA-H.CONTIG"
US-08-724-394A-20

Query Match 1.9%; Score 35.8; DB 2; Length 246240;
Best Local Similarity 52.3%; Pred. No. 28;
Matches 79; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

QY 1348 TTTAATTCATGACCTTTGGCTACCTGTGCGAGCTTTATAGACATTTCTTTTGA 1407
DB 85039 TTTAAGCCTTACAAATTTCTTTAAATTTCTGCTTAGCTTCAATTTTATATA 85098

QY 1408 GAAATATCAATAGATTTGGACCATGCTGCTTTAATGACTGTATTTGGGTATTCAT 1467
DB 85099 TAACTTTTATATATGTCATTTATACATCTTCATATGACTTATAGATTACATTT 85158

QY 1468 AAACATACCTTGATCATTTTGTATGCTATT 1498
DB 85159 TCCTAATATTAGATCTTTTGTATCATCTT 85189

RESULT 17
US-08-724-394A-21
Sequence 21, Application US/08724394A
Patent No. 5872237
GENERAL INFORMATION:
APPLICANT: Feder, John N.
APPLICANT: Krommal, Gregory S.
APPLICANT: Laufer, Peter M.
APPLICANT: Ruddy, David A.
APPLICANT: Thomas, Winston
APPLICANT: Tsuchihashi, Zenta
APPLICANT: Wolff, Roger K.
TITLE OF INVENTION: Megabase Transcript Map: No. 5872237el
TITLE OF INVENTION: Sequences and Antibodies Thereo
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP
STREET: Two Embarcadero Center, 8th Floor
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/724,394A
FILING DATE: 01-OCT-1996
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Filts, Renee A.
REGISTRATION NUMBER: 35,136
REFERENCE/DOCKET NUMBER: 017957-000100
TELECOMMUNICATION INFORMATION:

TELEPHONE: 415-576-0200
TELEFAX: 415-576-0300
INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 246240 base pairs
TYPE: nucleic acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: cdna
FEATURE:
NAME/KEY: misc feature
LOCATION: 1..246240
OTHER INFORMATION: /note= "HLA-H. CONTIG"
US-08-724-394A-21

Query Match 1.9%; Score 35.8; DB 2; Length 246240;
Best Local Similarity 52.3%; Pred. No. 28;
Matches 79; Conservative 0; Mismatches 72; Indels 0; Gaps 0;
QY 1348 TTTAATTCATGACCTTTGGTGTGCTGCGAGCTTTGTTTATAGCACATTTCTTTTGA 1407
DB 85039 TTTAAGCGCTACAATTTCTTAAATTCGCTTTAGCTTCATTTGTTATATAA 85098
QY 1408 GAAATTCGAATAGATTGCAAGCCATTTGGTGTGCTTTAATGACTGATTGGGGTATTCA 1467
DB 85099 TAACCTTTTATTATTGTTCCATCTCTATACATCTTCTAATGACTTATTAGTAGTTTACATT 85158
QY 1468 AAACAATCTTGTATCATTTTGTGATCTTTTGTATCTATT 1498
DB 85159 TCCTAATATTAGAAATCTTTTAGTCATCTT 85189

RESULT 18
US-08-724-394A-22
Sequence 22, Application US/08724394A
Patent No. 5872237
GENERAL INFORMATION:
APPLICANT: Feder, John N.
APPLICANT: Kronmal, Gregory S.
APPLICANT: Lauer, Peter M.
APPLICANT: Ruddy, David A.
APPLICANT: Thomas, Winston
APPLICANT: Tsuchihashi, Zenta
APPLICANT: Wolff, Roger K.
TITLE OF INVENTION: Megabase Transcript Map: No. 5872237el
TITLE OF INVENTION: Sequences and Antibodies Thereto
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP
STREET: Two Embarcadero Center, 8th Floor
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/724.394A
FILING DATE: 01-OCT-1996
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Fitts, Renee A.
REGISTRATION NUMBER: 35,136
REFERENCE/DOCKET NUMBER: 017957-000100
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-576-0200
TELEFAX: 415-576-0300
INFORMATION FOR SEQ ID NO: 22:
SEQUENCE CHARACTERISTICS:
LENGTH: 246240 base pairs

TYPE: nucleic acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: cdna
FEATURE:
NAME/KEY: misc feature
LOCATION: 1..246240
OTHER INFORMATION: /note= "HLA-H. CONTIG"
US-08-724-394A-22

Query Match 1.9%; Score 35.8; DB 2; Length 246240;
Best Local Similarity 52.3%; Pred. No. 28;
Matches 79; Conservative 0; Mismatches 72; Indels 0; Gaps 0;
QY 1348 TTTAATTCATGACCTTTGGTGTGCTGCGAGCTTTGTTTATAGCACATTTCTTTTGA 1407
DB 85039 TTTAAGCGCTACAATTTCTTAAATTCGCTTTAGCTTCATTTGTTATATAA 85098
QY 1408 GAAATTCGAATAGATTGCAAGCCATTTGGTGTGCTTTAATGACTGATTGGGGTATTCA 1467
DB 85099 TAACCTTTTATTATTGTTCCATCTCTATACATCTTCTAATGACTTATTAGTAGTTTACATT 85158
QY 1468 AAACAATCTTGTATCATTTTGTGATCTTTTGTATCTATT 1498
DB 85159 TCCTAATATTAGAAATCTTTTAGTCATCTT 85189

RESULT 19
US-08-487-826B-13
Sequence 13, Application US/08487826B
Patent No. 5993827
GENERAL INFORMATION:
APPLICANT: Sim, Kim L.
APPLICANT: Chitnis, Chetan
APPLICANT: Miller, Louis H.
APPLICANT: Peterson, David S.
APPLICANT: Su, Xin-zhaun
APPLICANT: Wellem, Thomas E.
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobbe Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/487.826B
FILING DATE: 10-SEP-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Israelsen, Ned
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER: NIH121.001CP1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-9550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 19124 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cdna
HYPOTHETICAL: NO
ANTI-SENSE: NO

US-08-487-826B-13

Query Match	1.9%;	Score 35.4;	DB 2;	Length 19124;
Best Local Similarity	44.7%;	Pred. No. 12;		
Matches 181;	Conservative	0;	Mismatches 221;	Indels 3;
			Gaps	1

Oy	1387	TATGACAACTTCTTTTGAAGAAAAATTCAAATAGATTGCAAGCCATGGCGCTTAT	1446
Db	3560	TATATATTTTTTTATACCTTGATCTCTACATGTTTTATATATATGATTTATTTAT	3619
Oy	1447	GACTGATTTGGGGATTCATTAACAATCTTGATCATTTTTGATCGTATTCGGAAGA	1506
Db	3620	TATTATACCTTATATATATATATATTTTTTACATTAGATATTAATATATGATCATCTATC	3679
Oy	1507	TGCGCAAGCAACTGTTTACCCTATGCAGATGTTTGTAGTAATGATGACCCCTCAAAAGC	1566
Db	3680	TATCATCTAT	3738
Oy	1567	GTTGACCGCGACGGTAATAGACAACAGCTACAACTCATACGTTTGTATGCTTTGTT	1626
Db	3740	ATTAGTAGATGATTTATATATATATACCTATTGGAAGAAATGAAACATATATATATTTAT	3798
Oy	1627	TATAGCGGCTCTCTGTCTTTAATTTTGCATTTATTTATGACCAATGAGATTTCTCTAG	1686
Db	3800	TATATGAACTTCATTTTTATTTGTTATATATATATATATATATATATATATATATATAT	3856
Oy	1687	AACTTATCGTCTCTTATATATGACACACCTGTGTGTTTATATGCTCCGTAAGAAA	1746
Db	3857	AATTTACACATGATATATATATATTTTATTTATTTGTGTTTATATATTTATATATATA	3916
Oy	1747	TGCGCAAAATAGACCGTAAACTTAATCTAACGTTGCAAT	1791
Db	3917	TGTGTTAATATATAGTGTATATATATTTTTTATATATATGACCAAT	3961

```

US-09-134-001C-2535/c
; Sequence 2535, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 2535
; LENGTH: 1245
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-2535
```

	Query Match	Similarity	Score	DB	Length	Matches	Conservative	Mismatches	Indels	Gaps
		51.3%	34.8	4	1245	81	0	77	0	0
Qy	569	GCCTTGGAGCCATTTTGTGATCTTAACCTTGACATACAGATATGGCCCTCAGACCTCTTTTC			628					
Db	311	GCACGGAAACCAAGTGTGATTACTATCTATCTGCGTAATGTGTTCTTGAATGTTC			252					
Qy	629	TTCTTGATACAGGCGCTATTAAAGGTTGCTTGACATGTATTTAGAAATTTCTCT			688					
Db	251	TTATTATCTTGATGTTCTATTACCTCGTAGAATAAAATTAAGCTTATTAATTAGAAC			192					
Qy	689	TCAATGTTAAGGCTCTTTTCATGACATCAATATTTTCTT			726					
Db	191	TCTATATGCTCTTGATTAATTAAGCCTTGATAGTCTT			154					

RESULT 21
ITE-09-869

```

US-08-869-696-1
Sequence 1, Application US/08869696C
Patent No. 603115
GENERAL INFORMATION
APPLICANT: Cameron-Mills, Verena
APPLICANT: Lok, Finn
APPLICANT: Sinzorgo, Catharina Maria Corneilia
APPLICANT: Van Den Dool, Ronald Taco Marinus
APPLICANT: Caspers, Martinus Petrus Maria
APPLICANT: Van Zeijl, Van Der Valk, Maria Joana
TITLE OF INVENTION: ARABINOXYLAN DEGRADATION
FILE REFERENCE: 11225-01US01
CURRENT APPLICATION NUMBER: US/08/869,696C
CURRENT FILING DATE: 1997-06-05
NUMBER OF SEQ ID NOS: 22
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1
LENGTH: 5529
TYPE: DNA
ORGANISM: barley
US-08-869-696-1

```

Query Match	1.9%;	Score	34.8;	DB	3;	Length	5529;
Best Local Similarity	58.8%;	Pred. No.	10;				
Matches	60;	Conservative	0;	Mismatches	42;	Indels	0;
						Gaps	0

Qy	1351	AATTCAGACCTTTTGCCCTACGTCGACGCTGTGTTATAGACAACTTTCTTTGAMAA	1410
Db	4320	AATTCAGATATTTTCTTACCATCAAGCTTTTGAAATGTATCAACTTTTCAAACA	4379
Qy	1411	AATTCATAATGATTTGCAACCCATGTCGCTTAATGACGT	1452
Db	4380	ATTCTGATTAATTTTAAACACATGCTTTTAAATTAATG	4421

RESULT 22
 US-08-961-527-126
 ; Sequence 126, Application US/08961527
 ; Patent No. 6420135
 ; GENERAL INFORMATION:
 ; APPLICANT: Charles Kunsch
 ; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
 ; NUMBER OF SEQUENCES: 391
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Human Genome Sciences, Inc.
 ; STREET: 9410 Key West Avenue
 ; CITY: Rockville
 ; STATE: Maryland
 ; COUNTRY: USA
 ; ZIP: 20850
 ;
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
 ; COMPUTER: HP Vectra 486/33
 ; OPERATING SYSTEM: MSDOS version 6.2
 ; SOFTWARE: ASCII Text
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/961,527
 ;
 ; FILING DATE:
 ; CLASSIFICATION: 424
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER:
 ;
 ; FILING DATE:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Brookes, A. Anders
 ; REGISTRATION NUMBER: 36,373
 ; REFERENCE/DOCKET NUMBER: PB340P1
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (301) 309-8504
 ; TELEFAX: (301) 309-8512
 ; INFORMATION FOR SEQ ID NO: 126:
 ; SEQUENCE CHARACTERISTICS:

US-08-757-653-175/c
; Sequence 175, Application US/08757653
; Patent No. 584369
; GENERAL INFORMATION:
; APPLICANT: Kaiser, Michael W.
; APPLICANT: Lyamichev, Victor I.
; APPLICANT: Lyamichev, Natasha
; TITLE OF INVENTION: Cleavage Of Nucleic Acid Using
; TITLE OF INVENTION: Thermostable FEN-1 Endonucleases
; NUMBER OF SEQUENCES: 190
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll, LLP
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States Of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/757,653
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Ingolia, Diane E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: FORS-02565
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 175:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1023 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1020
; US-08-757-653-175

Query Match 1.8%; Score 33.2; DB 2; Length 1023;
Best Local Similarity 49.4%; Pred. No. 13;
Matches 86; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

QY 336 TATATTATAGATTGGAGGCGTCATCGCTTGGAGCTGTTCTTGAATCTTTTGGCTTA 395
|||
DB 825 TATTGCATATATAATCCACATCGCTTGTCTTTGGAACTTGTCTAGCGGATCTTTTGAATG 766
|||
QY 396 TCTGGGAGCTCTACAGTATTGGATGGCCACTCACCTTGTGAGACTCGCTGGGATTTG 455
|||
DB 765 TCTAACATCTCTAAAGCTTTTAAAGCCTATGCGCTTATCTCTCTGGGTTGTATTC 706
|||
QY 456 TTCTGTATGAGGATGGCCGTAGATGCAAAATGTTCTTGTATTGAAAGAAATCC 509
|||
DB 705 TGTTCACAGAGGATTTGCTAGTTCAATGAGCTTTTCTCTGTAGACTTAAATTC 652
|||

RESULT 26
US-08-823-516-78/c
; Sequence 78, Application US/08823516
; Patent No. 5994069
; GENERAL INFORMATION:
; APPLICANT: Hall, Jeff G.
; APPLICANT: Lyamichev, Victor I.
; APPLICANT: Mast, Andrea L.
; APPLICANT: Brown, Mary Ann D.
; TITLE OF INVENTION: Detection Of Nucleic Acids By Multiple
; TITLE OF INVENTION: Sequential Invasive Cleavages

NUMBER OF SEQUENCES: 163
CORRESPONDENCE ADDRESS:
ADDRESSEE: Medlen & Carroll, LLP
STREET: 220 Montgomery Street, Suite 2200
CITY: San Francisco
STATE: California
COUNTRY: United States Of America
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/823,516
FILING DATE: 24-MAR-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US97/01072
FILING DATE: 21-JAN-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/759,038
FILING DATE: 02-DEC-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/758,314
FILING DATE: 02-DEC-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/756,386
FILING DATE: 29-NOV-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/682,853
FILING DATE: 12-JUL-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/599,491
FILING DATE: 24-JAN-1996
ATTORNEY/AGENT INFORMATION:
NAME: Ingolia, Diane E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: FORS-02736
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 78:
SEQUENCE CHARACTERISTICS:
LENGTH: 1023 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1020
US-08-823-516-78

Query Match 1.8%; Score 33.2; DB 2; Length 1023;
Best Local Similarity 49.4%; Pred. No. 13;
Matches 86; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

QY 336 TATATTATAGATTGGAGGCGTCATCGCTTGGAGCTGTTCTTGAATCTTTTGGCTTA 395
|||
DB 825 TATTGCATATATAATCCACATCGCTTGTCTTTGGAACTTGTCTAGCGGATCTTTTGAATG 766
|||
QY 396 TCTGGGAGCTCTACAGTATTGGATGGCCACTCACCTTGTGAGACTCGCTGGGATTTG 455
|||
DB 765 TCTAACATCTCTAAAGCTTTTAAAGCCTATGCGCTTATCTCTCTGGGTTGTATTC 706
|||
QY 456 TTCTGTATGAGGATGGCCGTAGATGCAAAATGTTCTTGTATTGAAAGAAATCC 509
|||
DB 705 TGTTCACAGAGGATTTGCTAGTTCAATGAGCTTTTCTCTGTAGACTTAAATTC 652
|||

RESULT 27
US-08-759-038-114/c

; Sequence 114, Application US/08759038
; Patent No. 6090543
; GENERAL INFORMATION:
; APPLICANT: Prudent, James R.
; APPLICANT: Hall, Jeff G.
; APPLICANT: Lyamichev, Victor I.
; APPLICANT: Brow, Mary Ann D.
; APPLICANT: Dahlberg, James E.
; TITLE OF INVENTION: Cleavage Of Nucleic Acids
; NUMBER OF SEQUENCES: 134
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll, LLP
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States Of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/759,038
; FILING DATE: 02-DEC-1996
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/
; FILING DATE: 29-NOV-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/682,853
; FILING DATE: 12-JUL-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/599,491
; FILING DATE: 24-JAN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Ingolia, Diane E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: FORS-02574
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 114:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1023 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1020
; US-08-759-038-114

Query Match 1.8%; Score 33.2; DB 3; Length 1023;
Best Local Similarity 49.4%; Pred. No. 13;
Matches 86; Conservative 0; Mismatches 88; Indels 0; Gaps 0;
QY 336 TATATTATAGATTGGAGCGCTCATCGCTTCGGAGCTGTTCTTCTGAATCTTTGCTTA 395
Db 825 TATTGCATATAAATCCACATCGCTTTGCTTTTGGAACTTGTCTAGCGGATCTTTGAGTG 766
QY 396 TCTGGCAGCTCTACAGTATTGGATGCGCCACTCACCTTGTGAGGACTCGCTGGGATTG 455
Db 765 TCTAACATCTCTAAAGCTTTTAAAGCCCTATGCCCTTTATTCCTCTCGGTTGATGTC 706
QY 456 TTCTTGCTATGGGATGCGCGTAGATGCAAAATGTTCTTGTTATTCGAAAGATCC 509
Db 705 TGTTCACACGAGGATTGCTAGTTCAATGAGCTTTTCTCTTGTTAGCTTTAATTC 652

RESULT 28
US-08-758-314-114/c

; Sequence 114, Application US/08758314
; Patent No. 6090606
; GENERAL INFORMATION:
; APPLICANT: Kaiser, Michael W.
; APPLICANT: Lyamichev, Victor I.
; APPLICANT: Lyamichev, Natasha
; TITLE OF INVENTION: Improved Cleavage Agents
; NUMBER OF SEQUENCES: 134
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll, LLP
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States Of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/758,314
; FILING DATE: 02-DEC-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/
; FILING DATE: 29-NOV-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/682,853
; FILING DATE: 12-JUL-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/599,491
; FILING DATE: 24-JAN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Ingolia, Diane E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: FORS-02575
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 114:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1023 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1020
; US-08-758-314-114

Query Match 1.8%; Score 33.2; DB 3; Length 1023;
Best Local Similarity 49.4%; Pred. No. 13;
Matches 86; Conservative 0; Mismatches 88; Indels 0; Gaps 0;
QY 336 TATATTATAGATTGGAGCGCTCATCGCTTCGGAGCTGTTCTTCTGAATCTTTGCTTA 395
Db 825 TATTGCATATAAATCCACATCGCTTTGCTTTTGGAACTTGTCTAGCGGATCTTTGAGTG 766
QY 396 TCTGGCAGCTCTACAGTATTGGATGCGCCACTCACCTTGTGAGGACTCGCTGGGATTG 455
Db 765 TCTAACATCTCTAAAGCTTTTAAAGCCCTATGCCCTTTATTCCTCTCGGTTGATGTC 706
QY 456 TTCTTGCTATGGGATGCGCGTAGATGCAAAATGTTCTTGTTATTCGAAAGATCC 509
Db 705 TGTTCACACGAGGATTGCTAGTTCAATGAGCTTTTCTCTTGTTAGCTTTAATTC 652

RESULT 29
US-08-844-274-11
; Sequence 11, Application US/08844274B
; Patent No. 6218185

Query Match	1.8%;	Score 33.2;	DB 4;	Length 5194;
Best Local Similarity	46.5%;	Pred. No. 27;		

RESULT 32
US-08-844-274-10/C
; Sequence 10, Application US/08844274H
; Patent No. 6218185
; GENERAL INFORMATION:
; APPLICANT: Frazer Jr., Malcom J.


```

; TITLE OF INVENTION: PiggyBac Transposon-Based Genetic Transformation System
; FILE REFERENCE: 0148.96
; CURRENT APPLICATION NUMBER: US/08/844,274B
; EARLIER FILING DATE: 1997-04-18
; EARLIER APPLICATION NUMBER: 60/016,234
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: Patentin Ver. 2.0 - beta
; SEQ ID NO 14
; LENGTH: 6723
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: piggyBac/opd
US-08-844-274-14

Query Match          1.8%; Score 33.2; DB 4; Length 6723;
Best Local Similarity 46.5%; Pred. No. 30; Mismatches 133; Indels 0; Gaps 0;
Matches 107; Conservative 0;

Qy 1313 TTGAATGCAATATGCTTCAGTCCGTATGCGCTTAATTCATGACCTTTGGCTACC 1372
    |||||
Db 5360 TTAAAGACAAATTAAGTATTTTGTAAAGAGAAAGTTAAAGTTTGTACT 5419

Qy 1373 TGTCAGCTTGTATTAGACAACTTTTGAAGAAATTCAAATGATTGGCAAGCC 1432
    |||||
Db 5420 TTAGAGAAATTTGAGTTTGTATTTTAAATTAATTAATTAATTAATTAATTTGT 5479

Qy 1433 ATTGCTCTTAATGACGTATGCGGTATCATTAACATCACTTGATCTTTGAT 1492
    |||||
Db 5480 TTGTTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 5539

Qy 1493 CGATTCGTAAGATCCGCAAGCAAGCTGTTTACCCCTTATGATGTTT 1542
    |||||
Db 5540 TTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 5589

RESULT 36
US-08-844-274-20
; Sequence 20, Application US/08844274B
; Patent No. 6218185
; GENERAL INFORMATION:
; APPLICANT: Fraser Jr., Malcolm J.
; APPLICANT: Shirk, Paul D.
; APPLICANT: Ellick, Teri A.
; APPLICANT: Pereira, Omahthage
; TITLE OF INVENTION: PiggyBac Transposon-Based Genetic Transformation System
; FILE REFERENCE: 0148.96
; CURRENT APPLICATION NUMBER: US/08/844,274B
; EARLIER FILING DATE: 1997-04-18
; EARLIER APPLICATION NUMBER: 60/016,234
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: Patentin Ver. 2.0 - beta
; SEQ ID NO 20
; LENGTH: 7560
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: p381.2hs/opd
US-08-844-274-20

Query Match          1.8%; Score 33.2; DB 4; Length 7560;
Best Local Similarity 46.5%; Pred. No. 32; Mismatches 133; Indels 0; Gaps 0;
Matches 107; Conservative 0;

Qy 1313 TTGAATGCAATATGCTTCAGTCCGTATGCGCTTAATTCATGACCTTTGGCTACC 1372
    |||||
Db 3798 TTAAAGACAAATTAAGTATTTTGTAAAGAGAAAGTTAAAGTTTGTACT 3857

Qy 1373 TGTCAGCTTGTATTAGACAACTTTTGAAGAAATTCAAATGATTGGCAAGCC 1432
```

```

Db 3858 TTGAAGAAATTTGACTTTTGTATTTTAAATTAATTAATTAATTAATTAATTTGT 3917

Qy 1433 ATTGCTCTTAATGACGTATGCGGTATTCATTAACATCACTTGATCTTTGAT 1492
    |||||
Db 3918 TTGTTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3977

Qy 1493 CGATTCGTAAGATCCGCAAGCAAGCTGTTTACCCCTTATGATGTTT 1542
    |||||
Db 3978 TTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 4027

RESULT 37
US-08-961-527-69/c
; Sequence 69, Application US/08961527
; Patent No. 6420135
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 391
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,527
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 69:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8898 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-961-527-69

Query Match          1.8%; Score 33.2; DB 4; Length 8898;
Best Local Similarity 50.0%; Pred. No. 34; Mismatches 83; Indels 0; Gaps 0;
Matches 83; Conservative 0;

Qy 1694 TCGCTCTTATATGACCAACCTGTTGTTATGTCGTAAGAAATGCTCA 1753
    |||||
Db 2656 TGGGACAGTTCAGAAAGACCGTACCGTATGCTGATTTTGGAAATGATATCTGA 2597

Qy 1754 AATTAAGTACCGTTAACTTAATCTTAACGTGACCAATATTAATCTCTTTGGACTT 1813
    |||||
Db 2596 AATATATGCTTCAAACTACTATTAAGAACACATAGTAATAATGAAATTTGAATTA 2537

Qy 1814 TAGTCCCAAGGCCCTGTGTATTAATTTATGACAAATTCAGT 1859
    |||||
Db 2536 TTCAATATTAATCTTATGCTTAATAAGCTGATGGAAGATTGAT 2491

RESULT 38
US-08-464-266-1/c
```

; Sequence 1, Application US/08464266
; Patent No. 5641652
; GENERAL INFORMATION:
; APPLICANT: ORO, Ph.D., ANTHONY E.
; APPLICANT: EVANS, Ph.D., RONALD M.
; TITLE OF INVENTION: INSECT RETINOID-LIKE RECEPTOR
; TITLE OF INVENTION: COMPOSITIONS AND METHODS
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark
; STREET: 444 South Flower Street, Suite 2000
; CITY: Los Angeles
; STATE: CA
; COUNTRY: United States
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/464,266
; FILING DATE: 05-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/013,975
; FILING DATE: 04-FEB-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/497,935
; FILING DATE: 22-FEB-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Reiter, Stephen E.
; REGISTRATION NUMBER: 31,192
; REFERENCE/DOCKET NUMBER: P41 9966
; TELEPHONE: 619-546-4737
; TELEFAX: 619-546-9392
; INFORMATION FOR SEQ ID NO: 1:
; LENGTH: 2304 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 163..1704
US-08-464-266-1

Query Match 1.8%; Score 33; DB 1; Length 2304;
Best Local Similarity 63.0%; Pred. No. 21;
Matches 51; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

QY 646 TATTAAAGGTTTCCTTGACATTTAGGAATTTCTCTTCAATGTTACGGCTCT 705
Db 2009 TAGTTAACTTTTACTTTTACTTGTGTCATTTCTCGTTCGTTTGTGTTCT 1950

QY 706 TTTCATGACTAAATTTTCTT 726
Db 1949 TTTCAGTTTGGATTGTT 1929

RESULT 39
US-08-464-272-1/c
; Sequence 1, Application US/08464272
; Patent No. 5688691
; GENERAL INFORMATION:
; APPLICANT: ORO, Ph.D., ANTHONY E.
; APPLICANT: EVANS, Ph.D., RONALD M.
; TITLE OF INVENTION: INSECT RETINOID-LIKE RECEPTOR
; TITLE OF INVENTION: COMPOSITIONS AND METHODS
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark
; STREET: 444 South Flower Street, Suite 2000
; CITY: Los Angeles
; STATE: CA
; COUNTRY: United States
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/464,272
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/013,975
; FILING DATE: 04-FEB-1993
; APPLICATION NUMBER: US 07/497,935
; FILING DATE: 22-FEB-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Reiter, Stephen E.
; REGISTRATION NUMBER: 31,192
; REFERENCE/DOCKET NUMBER: P41 9350
; TELEPHONE: 619-546-4737
; TELEFAX: 619-546-9392
; INFORMATION FOR SEQ ID NO: 1:
; LENGTH: 2304 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 163..1704
US-08-464-272-1

Query Match 1.8%; Score 33; DB 1; Length 2304;
Best Local Similarity 63.0%; Pred. No. 21;
Matches 51; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

QY 646 TATTAAAGGTTTCCTTGACATTTAGGAATTTCTCTTCAATGTTACGGCTCT 705
Db 2009 TAGTTAACTTTTACTTTTACTTGTGTCATTTCTCGTTCGTTTGTGTTCT 1950

QY 706 TTTCATGACTAAATTTTCTT 726
Db 1949 TTTCAGTTTGGATTGTT 1929

RESULT 40
US-08-464-514-1/c
; Sequence 1, Application US/08464514
; Patent No. 6265173
; GENERAL INFORMATION:
; APPLICANT: EVANS, RONALD M.
; APPLICANT: MCKEOWN, MICHAEL B.
; APPLICANT: ORO, ANTHONY E.
; APPLICANT: SEGRAVES, WILLIAM A.
; APPLICANT: YAO, TSO-PANG
; TITLE OF INVENTION: MULTIMERIC FORMS OF MEMBERS OF THE
; TITLE OF INVENTION: STEROID/THYROID SUPERFAMILY OF RECEPTORS WITH THE
; TITLE OF INVENTION: ULTRASPIRACLE RECEPTOR
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PRETTY, SCHROEDER, BRUEGGEMANN & CLARK
; STREET: 444 South Flower Street, Suite 2000
; CITY: Los Angeles
; STATE: California
; COUNTRY: United States
; ZIP: 90071

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/464,514
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/907,908
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Reiter, Stephen E.
REGISTRATION NUMBER: 31192
REFERENCE/DOCKET NUMBER: P41 9321
TELEPHONE: (619) 546-4737
TELEFAX: (619) 546-9392
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2304 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 163..1701
US-08-464-514-1

Query Match 1.8%; Score 33; DB 4; Length 2304;
Best Local Similarity 63.0%; Pred. No. 21;
Matches 51; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

Qy 646 TATTAAAGGTTTGCTTGACATTTTGAATTTTCTTCATGATGTTACGGCTCT 705
Db 2009 TAGTTTACCTTTTACTTTACTTCTGCTGTGCGATTTTCGTCGTTTGTTGTTCT 1950

Qy 706 TTTCATGACTAAATTTTCTT 726
Db 1949 TTTCAGTTTGGATTTTGT 1929

RESULT 41
US-08-486-403-1/c
Sequence 1, Application US/08486403
Patent No. 6281330
GENERAL INFORMATION:
APPLICANT: EVANS, RONALD M.
APPLICANT: MCKEOWN, MICHAEL B.
APPLICANT: ORO, ANTHONY E.
APPLICANT: SEGRAVES, WILLIAM A.
APPLICANT: YAO, TSO-PANG
TITLE OF INVENTION: MULTIMERIC FORMS OF MEMBERS OF THE
TITLE OF INVENTION: STEROID/THYROID SUPERFAMILY OF RECEPTORS WITH THE
TITLE OF INVENTION: ULTRASPIRACLE RECEPTOR
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: PRETTY, SCHROEDER, BRUEGGEMANN & CLARK
STREET: 444 South Flower Street, Suite 2000
CITY: Los Angeles
STATE: California
COUNTRY: United States
ZIP: 90071
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/486,403
FILING DATE:

CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/907,908
FILING DATE: 02-JUL-1992
ATTORNEY/AGENT INFORMATION:
NAME: Reiter, Stephen E.
REGISTRATION NUMBER: 31192
REFERENCE/DOCKET NUMBER: P41 9321
TELEPHONE: (619) 546-4737
TELEFAX: (619) 546-9392
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2304 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 163..1701
US-08-486-403-1

Query Match 1.8%; Score 33; DB 4; Length 2304;
Best Local Similarity 63.0%; Pred. No. 21;
Matches 51; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

Qy 646 TATTAAAGGTTTGCTTGACATTTTGAATTTTCTTCATGATGTTACGGCTCT 705
Db 2009 TAGTTTACCTTTTACTTTACTTCTGCTGTGCGATTTTCGTCGTTTGTTGTTCT 1950

Qy 706 TTTCATGACTAAATTTTCTT 726
Db 1949 TTTCAGTTTGGATTTTGT 1929

RESULT 42
US-09-149-476-221/c
Sequence 221, Application US/09149476
Patent No. 6420526
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: 186 Human Secreted proteins
FILE REFERENCE: P2002P1
CURRENT APPLICATION NUMBER: US/09/149,476
CURRENT FILING DATE: 1998-09-08
EARLIER APPLICATION NUMBER: PCT/US98/04493
EARLIER FILING DATE: 1998-03-06
EARLIER APPLICATION NUMBER: 60/040,162
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/040,333
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/038,621
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/040,626
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/040,334
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/040,336
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/040,163
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/047,600
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,615
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,597
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,502
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,633
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,583

EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,862
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,887
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,908
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/048,964
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/057,650
EARLIER FILING DATE: 1997-09-05
EARLIER APPLICATION NUMBER: 60/056,884
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/057,669
EARLIER FILING DATE: 1997-09-05
EARLIER APPLICATION NUMBER: 60/049,610
EARLIER FILING DATE: 1997-06-13
EARLIER APPLICATION NUMBER: 60/061,060
EARLIER FILING DATE: 1997-10-02

Query Match 1.7%; Score 32.6; DB 4; Length 504;
Best Local Similarity 51.4%; Pred. No. 14;
Matches 71; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

QY 1571 AGCCGACGGTAATGACACAGCTACACTATACAGTTTGTAAATGCTTTGTATATA 1630
DB 161 ATCCGTAAGTAATGAATGAATGAATGTAATCTGAAAGTTTAACTTTCTTTT 102
QY 1631 GGGCGCTCTGCTCTTAAATTTGCAATTTATTAATGACATAGGATCTTTAGAACT 1690
DB 101 TAGTTTTTTCTTTTGTGCAATTTTTTTTTTTTGGGGTTTGAGTGTGTTGAGGT 42
QY 1691 TTATGCTCTTTATATT 1708
DB 41 TTTTNNCCCTTTTTT 24

RESULT 43
US-09-008-979A-3
Sequence 3, Application US/0908979A
Patent No. 6080914

GENERAL INFORMATION:
APPLICANT: Comer, Timothy W.
TITLE OF INVENTION: Strawberry Promoters and Genes
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Carmen Rodriguez, Paralegal, Arnold, White & Durkee
STREET: P.O. Box 4433
CITY: Houston
STATE: Texas
COUNTRY: USA
ZIP: 77210-4433
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/008,979A
FILING DATE:
CLASSIFICATION: 800

ATTORNEY/AGENT INFORMATION:
NAME: Kammerer, Patricia A.
REGISTRATION NUMBER: 29,775
REFERENCE/DOCKET NUMBER: MOBT:058 (38-21(10613)A)
TELECOMMUNICATION INFORMATION:
TELEPHONE: (713) 787-1438
TELEFAX: (713) 789-2679
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1260 base pairs
TYPE: nucleic acid
STRANDEDNESS: double

TOPOLOGY: linear
US-09-008-979A-3

Query Match 1.7%; Score 32.4; DB 3; Length 1260;
Best Local Similarity 50.6%; Pred. No. 24;
Matches 78; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

QY 1390 AGCATTCTTTTGAAGAAATTCAAATAGATTTCAGAGCATGTGCTTAAATGAC 1449
DB 1069 ACCGATTGGAAGCGAGATATTCAGTCGATCGACCAAGTGTTCCTCCATTGGA 1128
QY 1450 TGTATGGGATTCATTAAACAATCTTTGATCATTTTATCGTATTCGAGATG 1509
DB 1129 TGCCGCTGGAATCCAAATGAACTGATTTTAAATTTTCTTAAAGAACCCCG 1188
QY 1510 CCAAGGAACCTTTTACCCCTATGATGTTTGA 1543
DB 1189 AGACTGCGCTGCTTGTGCTATGATCTTCA 1222

RESULT 44
US-09-460-618-3
Sequence 3, Application US/09460618
Patent No. 6235482

GENERAL INFORMATION:
APPLICANT: Comer, Timothy W.
TITLE OF INVENTION: Strawberry Promoters and Genes
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Carmen Rodriguez, Paralegal, Arnold, White & Durkee
STREET: P.O. Box 4433
CITY: Houston
STATE: Texas
COUNTRY: USA
ZIP: 77210-4433
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/460,618
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/008,979
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Kammerer, Patricia A.
REGISTRATION NUMBER: 29,775
REFERENCE/DOCKET NUMBER: MOBT:058 (38-21(10613)A)
TELECOMMUNICATION INFORMATION:
TELEPHONE: (713) 787-1438
TELEFAX: (713) 789-2679
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1260 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-09-460-618-3

Query Match 1.7%; Score 32.4; DB 4; Length 1260;
Best Local Similarity 50.6%; Pred. No. 24;
Matches 78; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

QY 1390 AGCATTCTTTTGAAGAAATTCAAATAGATTTCAGAGCATGTGCTTAAATGAC 1449
DB 1069 ACCGATTGGAAGCGAGATATTCAGTCGATCGACCAAGTGTTCCTCCATTGGA 1128
QY 1450 TGTATGGGATTCATTAAACAATCTTTGATCATTTTATCGTATTCGAGATG 1509
DB 1129 TGCCGCTGGAATCCAAATGAACTGATTTTAAATTTTCTTAAAGAACCCCG 1188

Qy 1510 CCAAGCGAACCTGTTTACCCCTATGCATGTTTTA 1543
Db 1189 AGACTGGCCTCTGCTTTGCTCTATGATCCTTCA 1222

RESULT 45

US-09-310-235B-3
; Sequence 3, Application US/09310235B
; Patent No. 6392030
; GENERAL INFORMATION:
; APPLICANT: Conner, Timothy
; TITLE OF INVENTION: STRAWBERRY PROMOTERS AND GENES
; FILE REFERENCE: MOBT058--1 (11899.0058.DVUS00)
; CURRENT APPLICATION NUMBER: US/09/310.235B
; CURRENT FILING DATE: 1998-05-12
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3
; LENGTH: 1260
; TYPE: DNA
; ORGANISM: fragaria x ananassa
US-09-310-235B-3

Query Match 1.7%; Score 32.4; DB 4; Length 1260;
Best Local Similarity 50.6%; Pred. No. 24;
Matches 78; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

Qy 1390 AGCATTTCCTTTTGAAGAAAATTCAAATAGATTGCAAGCCATTGGTGTCTTAAATGAC 1449
Db 1069 ACCGATTGGAAGCGAGCATATCCAGTCGATCAGCCAGTGTTCTCCCTCATTTGA 1128
Qy 1450 TGTATTGGGGTATTCAATTAACAATACTTTTGATCATTTTGTATTCGTATTCGTGAAGATCG 1509
Db 1129 TGCGTCTGGCATCCAATGAACGTATTTTAAATTTTCTTATAAAGAACCCCG 1188
Qy 1510 CCAAGCGAACCTGTTTACCCCTATGCATGTTTTA 1543
Db 1189 AGACTGGCCTCTGCTTTGCTCTATGATCCTTCA 1222

Search completed: December 17, 2002, 08:13:21
Job time : 1000 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (C) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 17, 2002, 09:05:16 ; Search time 1972 Seconds
(without alignments)
15308.513 Million cell updates/sec

Title: US-09-868-987-1
Perfect score: 1864
Sequence: 1 atggacttcgcattgtc.....atgacaattcagataatgc 1864

Scoring table: OLIGO NUC
Gapop 60.0 , Gapext 60.0

Searched: 16154066 seqs, 8097743376 residues

Word size : 0

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database :

EST:*

1: em_estba:*

2: em_esthum:*

3: em_estin:*

4: em_estmu:*

5: em_estov:*

6: em_estpl:*

7: em_estro:*

8: em_hcc:*

9: gb_est1:*

10: gb_est2:*

11: gb_hcc:*

12: gb_est3:*

13: gb_est4:*

14: gb_est5:*

15: em_estfun:*

16: em_estom:*

17: gb_gss:*

18: em_gss_hum:*

19: em_gss_inv:*

20: em_gss_pln:*

21: em_gss_vrt:*

22: em_gss_fun:*

23: em_gss_mam:*

24: em_gss_mus:*

25: em_gss_Other:*

26: em_gss_pro:*

27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	22	1.2	595	14	BQ038573
C 2	22	1.2	614	17	BH616209
C 3	22	1.2	647	9	AJ396413
C 4	21	1.1	242	17	AZ781921
C 5	21	1.1	533	10	BB283826
C 6	20	1.1	171	9	A1637310

C 80	19	1.0	606	9	AI993818	701514781	C 153	18	1.0	169	9	AUI85971	AUI85971
81	19	1.0	606	9	AA222578	my16e05.r	154	18	1.0	192	10	BE051996	BE051996
82	19	1.0	610	12	BF518890	EST456422	155	18	1.0	198	9	AV163828	AV163828
C 83	19	1.0	611	17	AZ61868	IM0267J21	156	18	1.0	207	17	AZ575764	AZ575764
84	19	1.0	624	17	AQ038180	CITBI-EI-	157	18	1.0	220	13	BI121623	BI121623
85	19	1.0	632	17	AQ379336	REC11-16	158	18	1.0	223	12	BE856208	BE856208
C 86	19	1.0	639	17	AG141283	Pan trogl	159	18	1.0	225	9	AUI257097	AUI257097
87	19	1.0	643	17	AZ573164	314PvD02	160	18	1.0	228	9	AI827328	AI827328
C 88	19	1.0	653	17	AZ259849	RPCI-23-1	161	18	1.0	229	12	BG013808	BG013808
C 89	19	1.0	665	17	AQ480736	RPCI-11-2	162	18	1.0	237	9	AV218287	AV218287
C 90	19	1.0	667	10	BE536878	601064885	163	18	1.0	246	9	AA366819	AA366819
C 91	19	1.0	677	12	BF714797	mab03d03-	164	18	1.0	248	12	BE986213	BE986213
92	19	1.0	678	10	BE264998	601193827	165	18	1.0	253	10	BE068699	BE068699
93	19	1.0	685	9	AL503713	AL503713	166	18	1.0	264	12	BE984790	BE984790
94	19	1.0	686	14	BQ597373	MI-P-A2-a	167	18	1.0	267	9	AV094338	AV094338
95	19	1.0	692	12	BF971071	602270624	168	18	1.0	269	12	BF326607	BF326607
C 96	19	1.0	694	17	AZ199764	SP 1040 A	169	18	1.0	271	9	AV092510	AV092510
C 97	19	1.0	695	17	AZ540662	226PvG07	170	18	1.0	271	9	AV141665	AV141665
C 98	19	1.0	713	17	BH292163	CH230-186	171	18	1.0	272	10	AW933803	AW933803
C 99	19	1.0	714	17	AG009030	Homo babi	172	18	1.0	275	9	AJ500448	AJ500448
100	19	1.0	728	12	BG697501	60260695	173	18	1.0	276	9	AJ500137	AJ500137
101	19	1.0	731	13	BI755220	603023227	174	18	1.0	281	9	AA484891	AA484891
102	19	1.0	733	17	BH181011	017 H 06-	175	18	1.0	281	10	BB022698	BB022698
103	19	1.0	733	17	CNS07MCP	AL617961 T7 end of	176	18	1.0	281	14	BO696513	BO696513
104	19	1.0	741	17	BH523054	HOGJ167TR	177	18	1.0	289	17	AZ252904	AZ252904
105	19	1.0	743	12	BG715477	602675447	178	18	1.0	293	9	AA123201	AA123201
C 106	19	1.0	751	9	AUI25363	AUI25363	179	18	1.0	294	14	Z14535	Z14535
107	19	1.0	764	13	BI758569	603023228	180	18	1.0	297	10	AV817576	AV817576
C 108	19	1.0	771	13	BI753921	603022928	181	18	1.0	299	9	AA921709	AA921709
C 109	19	1.0	780	9	AUI17917	AUI17917	182	18	1.0	310	14	W56262	W56262
110	19	1.0	790	12	BG699238	602678909	183	18	1.0	311	10	AI014820	AI014820
111	19	1.0	796	12	BG699445	602679138	184	18	1.0	316	9	AI702276	AI702276
C 112	19	1.0	806	9	AUI117827	AUI117827	185	18	1.0	319	10	AV932928	AV932928
113	19	1.0	807	13	BI648132	603278450	186	18	1.0	320	9	AI511305	AI511305
114	19	1.0	812	9	AUI43540	AUI43540	187	18	1.0	321	9	AV009727	AV009727
C 115	19	1.0	815	9	AUI26380	AUI26380	188	18	1.0	329	12	BE221229	BE221229
116	19	1.0	827	9	AUI26380	AUI26380	189	18	1.0	330	14	D74236	D74236
117	19	1.0	833	12	BG704268	602687643	190	18	1.0	331	10	AW933706	AW933706
118	19	1.0	842	12	BG697733	602661209	191	18	1.0	332	10	BB315355	BB315355
119	19	1.0	845	12	BG870872	602792283	192	18	1.0	333	10	BE646309	BE646309
120	19	1.0	862	12	BG714892	602675746	193	18	1.0	333	12	BF447278	BF447278
C 121	19	1.0	872	9	AUI32302	AUI32302	194	18	1.0	340	10	AW151093	AW151093
122	19	1.0	873	12	BE912645	601663688	195	18	1.0	340	10	AW279831	AW279831
123	19	1.0	881	12	BE788310	601480113	196	18	1.0	341	9	AUI10627	AUI10627
124	19	1.0	883	14	BQ961422	AGENCOURT	197	18	1.0	342	12	BG225557	BG225557
125	19	1.0	886	12	BG765535	602739450	198	18	1.0	344	14	T69881	T69881
126	19	1.0	890	14	BQ226575	AGENCOURT	199	18	1.0	350	9	AUI11156	AUI11156
127	19	1.0	903	13	BI560672	603254701	200	18	1.0	352	9	AA930406	AA930406
128	19	1.0	905	17	AZ543188	EMTGF34TR	201	18	1.0	352	9	AA930406	AA930406
129	19	1.0	912	12	BG031637	602229785	202	18	1.0	360	9	AJ467982	AJ467982
C 130	19	1.0	916	12	BF103352	BI103352	203	18	1.0	360	9	AJ468982	AJ468982
131	19	1.0	951	13	BI462668	603202109	204	18	1.0	360	9	AV189465	AV189465
132	19	1.0	957	12	BG036197	602326945	205	18	1.0	360	14	C09172	C09172
133	19	1.0	969	9	AL514362	AL514362	206	18	1.0	360	14	C40460	C40460
134	19	1.0	978	9	AL545110	AL545110	207	18	1.0	360	14	C40830	C40830
135	19	1.0	984	13	BI687590	603312552	208	18	1.0	360	14	C42115	C42115
136	19	1.0	1062	10	BE617480	601442138	209	18	1.0	360	14	C43462	C43462
137	19	1.0	1062	12	CNS03908	602325044	210	18	1.0	360	14	D67729	D67729
138	19	1.0	1075	12	BG035417	602325044	211	18	1.0	360	14	D67729	D67729
139	19	1.0	1086	14	BM924516	AGENCOURT	212	18	1.0	360	14	D70639	D70639
140	19	1.0	1088	13	BM477275	AGENCOURT	213	18	1.0	360	14	D74316	D74316
141	19	1.0	1098	14	BQ421029	AGENCOURT	214	18	1.0	360	14	D74393	D74393
142	19	1.0	1114	13	BM462885	AGENCOURT	215	18	1.0	360	14	D74404	D74404
143	19	1.0	1142	13	BM547681	AGENCOURT	216	18	1.0	360	14	D74759	D74759
144	19	1.0	1144	13	BM547681	AGENCOURT	217	18	1.0	360	14	D75017	D75017
145	19	1.0	1692	13	BM416552	CE21649 M	218	18	1.0	360	14	D75884	D75884
C 146	19	1.0	1890	13	BM416552	CE21649 M	219	18	1.0	360	14	D75957	D75957
147	19	1.0	2646	11	BC004332	Homo sapi	220	18	1.0	361	9	AI974743	AI974743
C 148	19	1.0	42	17	TA255A05P	AL482885 T1 drucei	221	18	1.0	365	17	AZ886610	AZ886610
C 149	19	1.0	92	17	AL766294	AL766294 Arabidops	222	18	1.0	366	10	BE496520	BE496520
150	19	1.0	129	17	BH454374	BOG6S25TR	223	18	1.0	371	10	BE209089	BE209089
C 151	19	1.0	138	12	BF641717	NF066C08I	224	18	1.0	374	12	BF115459	BF115459
152	19	1.0	143	17	AO471195	CITBI-EI-	225	18	1.0	375	10	BE008957	BE008957
C 152	19	1.0	168	13	BI189418	e4ff11fs.f	225	18	1.0	375	10	BE008957	BE008957

C 810	17	0.9	253	14	R88637	R88637 ym95e05.r1	C 883	17	0.9	292	14	BQ456708	BQ456708
C 811	17	0.9	255	9	AV249135	AV249135 AV249135	C 884	17	0.9	294	9	AL674235	AL674235
C 812	17	0.9	255	14	BQ455293	BQ455293 ke1lg08.y	C 885	17	0.9	295	12	BG213069	BG213069
C 813	17	0.9	255	14	BQ455994	BQ455994 ke25b05.y	C 886	17	0.9	297	9	AA554193	AA554193
C 814	17	0.9	255	14	BQ456165	BQ456165 ke27c05.y	C 887	17	0.9	298	13	BI883127	BI883127
C 815	17	0.9	256	9	AV173835	AV173835 AV173835	C 888	17	0.9	299	12	BG885482	BG885482
C 816	17	0.9	256	14	BQ455565	BQ455565 ke20b12.y	C 889	17	0.9	299	13	BI399057	BI399057
C 817	17	0.9	257	14	BQ454949	BQ454949 ke16f01.y	C 890	17	0.9	299	14	BQ455495	BQ455495
C 818	17	0.9	257	14	BQ455514	BQ455514 ke14h03.y	C 891	17	0.9	300	10	AV179112	AV179112
C 819	17	0.9	257	14	BQ455690	BQ455690 ke21f06.y	C 892	17	0.9	300	10	AW408974	AW408974
C 820	17	0.9	257	14	BQ455825	BQ455825 ke23c04.y	C 893	17	0.9	300	14	C30455	C30455
C 821	17	0.9	257	14	BQ456523	BQ456523 ke39h02.y	C 894	17	0.9	300	14	C545187	C545187
C 822	17	0.9	257	14	BQ457166	BQ457166 ke35f12.y	C 895	17	0.9	301	14	BQ454959	BQ454959
C 823	17	0.9	257	14	N72880	N72880 za04c01.r1	C 896	17	0.9	302	9	AV166893	AV166893
C 824	17	0.9	258	10	B8278832	B8278832 B8278832	C 897	17	0.9	303	9	AI679701	AI679701
C 825	17	0.9	261	9	AV241140	AV241140 AV241140	C 898	17	0.9	304	14	BQ456071	BQ456071
C 826	17	0.9	261	17	AZ378543	AZ378543 1M0133C20	C 899	17	0.9	305	9	AV067751	AV067751
C 827	17	0.9	262	9	AA554469	AA554469 nk30g07.s	C 900	17	0.9	305	9	AV166180	AV166180
C 828	17	0.9	262	10	BB609199	BB609199 BB609199	C 901	17	0.9	305	10	B8042937	B8042937
C 829	17	0.9	264	10	B8573114	B8573114 B8573114	C 902	17	0.9	305	12	BF092641	BF092641
C 830	17	0.9	264	14	BU021548	BU021548 QHE30J12.	C 903	17	0.9	305	12	BF369425	BF369425
C 831	17	0.9	265	9	AA811873	AA811873 OB40f06.s	C 904	17	0.9	305	14	BQ846868	BQ846868
C 832	17	0.9	265	9	AI400706	AI400706 tg93f12.x	C 905	17	0.9	306	10	BE505633	BE505633
C 833	17	0.9	265	10	AW184795	AW184795 se82f06.y	C 906	17	0.9	307	12	BF093402	BF093402
C 834	17	0.9	265	14	T40425	T40425 ya33h12.r1	C 907	17	0.9	307	14	R18136	R18136
C 835	17	0.9	266	14	BM828938	BM828938 K-EST0101	C 908	17	0.9	308	10	BB061728	BB061728
C 836	17	0.9	267	10	BB586015	BB586015 BB586015	C 909	17	0.9	308	17	AZ6311265	AZ6311265
C 837	17	0.9	267	14	BQ845261	BQ845261 QGA16D22.	C 910	17	0.9	308	17	AW518598	AW518598
C 838	17	0.9	267	14	BQ854197	BQ854197 QGB22J16.	C 911	17	0.9	309	10	AW518598	AW518598
C 839	17	0.9	267	14	BQ864569	BQ864569 QGC27B03.	C 912	17	0.9	311	10	B8021242	B8021242
C 840	17	0.9	267	14	BQ875683	BQ875683 QGI8M21.y	C 913	17	0.9	313	14	D11665	D11665
C 841	17	0.9	267	14	BQ895861	BQ895861 QGE7P02.y	C 914	17	0.9	313	14	D11673	D11673
C 842	17	0.9	268	14	BQ456505	BQ456505 ke39f04.y	C 915	17	0.9	314	9	AA039257	AA039257
C 843	17	0.9	268	14	BQ456641	BQ456641 ke41d01.y	C 916	17	0.9	314	10	BB828236	BB828236
C 844	17	0.9	268	14	BQ456752	BQ456752 ke42f07.y	C 917	17	0.9	316	9	AI224987	AI224987
C 845	17	0.9	268	14	BQ457224	BQ457224 ke36e03.y	C 918	17	0.9	316	9	AI826482	AI826482
C 846	17	0.9	268	14	BQ457396	BQ457396 ke38e07.y	C 919	17	0.9	316	9	AA152346	AA152346
C 847	17	0.9	269	14	BQ457035	BQ457035 ke34a09.y	C 920	17	0.9	316	13	BM030148	BM030148
C 848	17	0.9	269	14	BQ457362	BQ457362 ke38b06.y	C 921	17	0.9	318	10	AV525633	AV525633
C 849	17	0.9	271	10	AV409714	AV409714 AV409714	C 922	17	0.9	317	17	AZ870324	AZ870324
C 850	17	0.9	271	10	AW626800	AW626800 pa19f06.y	C 923	17	0.9	319	9	AI189729	AI189729
C 851	17	0.9	271	10	BB494474	BB494474 BB494474	C 924	17	0.9	319	9	AI430164	AI430164
C 852	17	0.9	272	10	AV426778	AV426778 AV426778	C 925	17	0.9	319	14	Z44306	Z44306
C 853	17	0.9	272	14	BQ455397	BQ455397 ke13c09.y	C 926	17	0.9	322	9	AA648351	AA648351
C 854	17	0.9	272	14	BQ848360	BQ848360 QGA6P18.y	C 927	17	0.9	323	9	AU264082	AU264082
C 855	17	0.9	273	14	AV066352	AV066352 AV066352	C 928	17	0.9	323	12	BG462241	BG462241
C 856	17	0.9	273	10	BB067670	BB067670 BB067670	C 929	17	0.9	323	17	AQ241399	AQ241399
C 857	17	0.9	273	10	BE658326	BE658326 GM700005B	C 930	17	0.9	324	9	AI991373	AI991373
C 858	17	0.9	273	14	BQ457176	BQ457176 ke35h03.y	C 931	17	0.9	324	9	AA469506	AA469506
C 859	17	0.9	275	10	BB254529	BB254529 BB254529	C 932	17	0.9	324	14	Z40956	Z40956
C 860	17	0.9	275	10	BE243761	BE243761 TCBAPID04	C 933	17	0.9	326	9	AV051686	AV051686
C 861	17	0.9	277	14	BQ863142	BQ863142 QGC23B08.	C 934	17	0.9	326	10	AV408234	AV408234
C 862	17	0.9	278	9	AV086502	AV086502 AV086502	C 935	17	0.9	327	14	F10916	F10916
C 863	17	0.9	278	9	AV135840	AV135840 AV135840	C 936	17	0.9	328	9	AI882653	AI882653
C 864	17	0.9	278	14	BQ457180	BQ457180 ke35h07.y	C 937	17	0.9	329	9	AA614685	AA614685
C 865	17	0.9	279	9	AI840771	AI840771 UI-M-AW0-	C 938	17	0.9	329	12	BF813846	BF813846
C 866	17	0.9	281	14	BQ858042	BQ858042 QGB9C24.y	C 939	17	0.9	329	17	AQ365571	AQ365571
C 867	17	0.9	281	14	T40263	T40263 ya31b04.r4	C 940	17	0.9	331	9	AL727382	AL727382
C 868	17	0.9	282	10	BB456001	BB456001 BB456001	C 941	17	0.9	331	17	BH643898	BH643898
C 869	17	0.9	282	14	BQ818898	BQ818898 QGC19C05.r	C 942	17	0.9	334	9	AI205212	AI205212
C 870	17	0.9	283	9	AA500668	AA500668 vg10d05.x	C 943	17	0.9	334	9	AI944870	AI944870
C 871	17	0.9	283	14	BQ997771	BQ997771 QGI17F23.	C 944	17	0.9	334	10	BB315308	BB315308
C 872	17	0.9	283	17	CNS07TF2	AL626624 T7 end of	C 945	17	0.9	334	14	BQ455636	BQ455636
C 873	17	0.9	285	9	AA610417	AA610417 np91b03.s	C 946	17	0.9	334	14	N63615	N63615
C 874	17	0.9	285	12	BG188719	BG188719 RST7747.A	C 947	17	0.9	335	12	BE974984	BE974984
C 875	17	0.9	288	10	BB378779	BB378779 BB378779	C 948	17	0.9	337	9	AI244781	AI244781
C 876	17	0.9	289	10	BB572332	BB572332 BB572332	C 949	17	0.9	337	14	F09524	F09524
C 877	17	0.9	291	9	AI869917	AI869917 we75h01.x	C 950	17	0.9	337	17	AZ034580	AZ034580
C 878	17	0.9	292	10	BB024579	BB024579 BB024579	C 951	17	0.9	338	9	AA907359	AA907359
C 879	17	0.9	292	10	BB068318	BB068318 BB068318	C 952	17	0.9	339	10	AW492329	AW492329
C 880	17	0.9	292	12	BG201423	BG201423 RST20888	C 953	17	0.9	339	10	AW551661	AW551661
C 881	17	0.9	292	12	BE717924	BE717924 MK2-HT078	C 954	17	0.9	339	17	BH418319	BH418319
C 882	17	0.9	292	14	BQ357280	BQ357280 IL5-HT070	C 955	17	0.9	340	9	AA730983	AA730983

956	17	0.9	341	10	BE575543
957	17	0.9	342	10	AV954787
958	17	0.9	342	10	AM289448
C 959	17	0.9	343	9	A1263127
C 960	17	0.9	343	10	AM093948
C 961	17	0.9	345	9	AA687187
C 962	17	0.9	345	9	A143657
C 963	17	0.9	345	13	T07776
964	17	0.9	346	13	BM168553
965	17	0.9	347	14	C90429
966	17	0.9	349	9	AA934410
C 967	17	0.9	349	9	A1212731
C 968	17	0.9	349	9	A1944550
C 969	17	0.9	350	9	A1158187
C 970	17	0.9	350	10	AM197747
C 971	17	0.9	350	14	BM655794
C 972	17	0.9	352	17	AZ070442
973	17	0.9	353	10	AW706656
974	17	0.9	355	10	AM439949
C 975	17	0.9	355	14	H485661
C 976	17	0.9	359	10	HA886694
C 977	17	0.9	359	12	BF221984
C 978	17	0.9	359	13	BM184924
C 979	17	0.9	360	9	AA813897
980	17	0.9	360	14	R15202
981	17	0.9	360	17	AZ891026
C 982	17	0.9	361	9	AA690062
C 983	17	0.9	361	13	B1496085
C 984	17	0.9	362	9	AA981996
C 985	17	0.9	362	9	AL804921
C 986	17	0.9	362	9	AA425420
C 987	17	0.9	362	10	AW415185
C 988	17	0.9	362	12	BF115345
C 989	17	0.9	363	14	BQ844416
990	17	0.9	363	12	AM238471
C 991	17	0.9	363	17	AO634984
C 992	17	0.9	364	9	A1372036
C 993	17	0.9	364	10	BE500966
C 994	17	0.9	365	12	BG000124
C 995	17	0.9	365	14	H78321
C 996	17	0.9	366	9	A1004695
C 997	17	0.9	367	9	AA492434
998	17	0.9	367	13	AV980531
999	17	0.9	367	10	B1924752
1000	17	0.9	368	9	A1270597
					BE575543 dc58d0d2.x
					AV954787 AV954787
					AM289448 NXNV001D1
					A1263127 q980a03.x
					AM093948 EST287128
					AA687187 n2u6d08.s
					A143657 q721a09.x
					T07776 EST05666.Fe
					BM168523 EST5710466
					C90429 C90429.Dict
					AA934410 C9049C06.s
					A1212731 x9g1a1a.t
					A1944550 b5o3g08.y
					A1158187 ud25h01.r
					AM197747 x188a03.x
					BM655794 mgc8003XMX
					AZ070442 RPTC1-23-4.r
					AW706656 BK01b09.y
					AM439949 ha11f11.x
					H485661 yq78g05.s1
					HA886694 RCl-OT008
					BF221984 7P40d04.x
					BM184924 f84d12.y
					AA813897 ny95c09.s
					R15202 yf88a10.r1
					AZ891026 RPTC1-24-2
					AA690062 v178d04.s
					B1496085 df122f12
					AA981996 ua34g02.r
					AL804921 AL804921
					AA425420 zw44e11.s
					AW415185 44298.MA
					BF115345 7m89f12.x
					BQ844416 QQa13M23
					AM238471 xp26d05.x
					AO634984 RPTC1-11-4
					A1372036 ta3sh02.x

ALIGNMENTS

RESULT 1	LOCUS	DEFINITION
BO038573	595 bp	linear EST 01-MAY-2002
gnic.pfk09	m19 normalized chicken lymphoid cDNA library	Gallus
gallus cDNA clone	gnic.pfk09 m19 5'	similar to sp 042242 SMO_CHICK
smoothened homolg	precursor (SMO) gb AA84389.1	(AF019577)
smoothened (Gallus gallus)	mRNA sequence.	

ACCESSION	BQ038573
VERSION	BQ038573.2
KEYWORDS	EST.
SOURCE	chicken.

ORGANISM

REFERENCE 1 (bases 1 to 595)

AUTHORS
Morgan, R. W. and Bull

TITLE Chicken Lymphoid Esti

On Mar 27, 2002 this sequence replaced gi:19772113

Molecular Endocrinology
University of Delaware

40 Townsend Hall, Newark, DE 19717, USA
Tel: 302 831-1345
Fax: 302-831-3411
Email: joan@Del.Edu, www.chickest.udel.edu
Location/Qualifiers
source 1..595

```

/clone lib="normalized chicken lymphoid cdna library"
/sex="Male and Female"
/tissue_type="thymus, bursa, spleen, PBL, bone marrow"
/lab_host="E.coli EMD108"
/motc="Vector: pCMVSPORT 6"

```

Query Match	1.2%	Score 22;	DB 14;	Length 595;
Best Local Similarity	100.0%;	Pred. No. 14;		
Matches	22;	Conservative	0;	Mismatches
			0;	Indels
				Gaps

Qy	1277	GCTTGGCAATCATCTTGCTCT	1298
Db	462	GCTTGGCAATCATCTTGCTCT	441

RESULT 2	614 bp	DNA	linear	GSS 28-JAN-2007
BH616209	BH616209			
LOCUS				
DEFINITION	BMAAC307A04SP6_P5U Brugia malayi Genomic Bac Library 3 Brugia			
	malaya genomic, DNA sequence.			

KEYWORDS	GSS.
SOURCE	Brugia malayi.
ORGANISM	Brugia malayi.

Eurytota; Metacoa; Nematoa; Cnidozoa; Splanchna; Ectozoa; Onchocercidae; Brugia.

REFERENCE
AUTHORS
1. Coates J. C. 1971
Whitton, C., Daub, J., Ware, J., Quail, M., Hall, N., Foster,
J., Guiliano, D., Slacko, B. and Blaxter, M.

TITLE	Genome survey sequences from the human parasitic nematode <i>Brugia malayi</i>
JOURNAL	Unpublished (2000)
COMMENT	Contact: Blaxter ML [email: b.laxter@glasgow.ac.uk] and Poulston BJ [email: j.poulston@glasgow.ac.uk]

Institute of Cell, Animal and Population Biology
University of Edinburgh
Ashworth Labs, King's Buildings, West Mains Road, Edinburgh, EH9 1JH

3JT, UK
Tel: +44 131 650 6760
Fax: +44 131 670 5450

Email: mark.blaxter@ed.ac.uk
 Sequenced from the Brugia malayi BAC library constructed by Claire

Whitton and Dr Mike Quail. The sequence was generated by the Pathogen Sequencing Unit, The Sanger Institute, Cambridge, UK in collaboration with Mark Blaxter, ICAPB, University of Edinburgh.

Edinburgh, UK.
Seq primer: SP6 (ATTAGTGACACTATAG

```

FEATURES      Location/Qualifiers
source        1. .614

```

```
/organism="Bruglia malayi"  
/strain="TRS"  
/db xref="taxon:6279"
```

```

/clone_lib="Brugia malayi Genomic Bac Library 3"
/sex="Mixed (male and female)"

```

```

/tissue_type="mole parasite"
/dev_stage="microfilaria (L1)"
/notes="Vector: PBACE3.6; Site 1: Bamh I; Brugia malay."

```

genomic DNA was partially cleaved with *Sau3A* I and size fractionated. 7,392 clones were generated with mean insert size of 48 bp. The library was constructed by Claire

size 100 kbp. The library was constructed by Genoscreen, Inc. and screened by Dr. Whitton, Blaxter Nematode Genetics Lab, University of

Edinburgh, UK, and Dr Mike Quail, The Pathogen Sequencing Unit, the Sanger Centre, Cambridge, UK."

BASE COUNT 187 a 108 c 103 g 216 t
ORIGIN

Query Match 1.2%; Score 22; DB 17; Length 614;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1392 CACATTTCTTTTGAAGAAAT 1413
|||||
Db 217 CACATTTCTTTTGAAGAAAT 238

RESULT 3
AJ396413/c
LOCUS
DEFINITION AJ396413 dkfz426 Gallus gallus cDNA clone 26f2r1, mRNA sequence.
ACCESSION AJ396413
VERSION AJ396413.1 GI:7128327
KEYWORDS EST.
SOURCE chicken.

ORGANISM Gallus gallus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.

REFERENCE 1 (bases 1 to 647)
AUTHORS Abdrahmanov, I., Lodygin, D., Gerth, P., Arakawa, H., Law, A., Plachy

TITLE A large database of chicken bursal ESTs as a resource for the

JOURNAL analysis of vertebrate gene function
MEDLINE Genome Res. 10 (12), 2062-2069 (2000)
COMMENT 20568495

FEATURES
source
1. 647
/organism="Gallus gallus"
/strain="CB"
/db_xref="taxon:9031"
/clone="26f2r1"
/clone_lib="dkfz426"
/tissue_type="Bursa of Fabricius"
/cell_type="Bursal lymphocyte"
/dev_stage="2-3 weeks old"
BASE COUNT 137 a 187 c 191 g 132 t
ORIGIN

Query Match 1.2%; Score 22; DB 9; Length 647;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1277 GCTTTGGCAATCATCTTCTCT 1298
|||||
Db 435 GCTTTGGCAATCATCTTCTCT 414

RESULT 4
AZ781921
LOCUS
DEFINITION 2M0021L08R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
clone UUGC2M0021L08 R, DNA sequence.
ACCESSION AZ781921
VERSION AZ781921.1 GI:12915097
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 242)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

AUTHORS

Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,
Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly,
M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A.
and Wright, D., Weiss, R.

Mouse whole genome scaffolding with paired end reads from 10kb

TITLE

plasmid inserts

JOURNAL

COMMENT Unpublished (2000)

Contact: Robert B. Weiss

University of Utah Genome Center

Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT

84112, USA

Tel: 801 585 5606

Fax: 801 585 7177

Email: ddunn@genetics.utah.edu

Insert Length: 10000 Std Error: 0.00

Plate: 0021 row: L column: 08

Seq primer: CACACAGGAACACGCTATGACC

Class: plasmid ends

High quality sequence stop: 242.

Location/Qualifiers

1. 242

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="UUGC2M0021L08"

/clone_lib="Mouse 10kb plasmid UUGC1M library"

/sex="Male"

/lab_host="E. Coli strain XL10-Gold, TI-resistant, F-"

/notes="Vector: PWD42nv; Purified genomic DNA from M.

musculus C57BL/6J (male) was obtained from the Jackson

Laboratory Mouse DNA Resource

(http://www.jax.org/resources/documents/dnares/). The DNA

was hydrodynamically sheared by repeated passage through a

0.005 inch orifice at constant velocity. The sheared DNA

polynucleotide kinase. Adapter oligonucleotides were

ligated to the blunt ends in high molar excess. The

adapted DNA was purified and size-selected for a 9.5 to

10.5 kb range using preparative agarose gel

electrophoresis. Vector DNA was prepared from a derivative

of PWD42 (gl|4732114|gb|AF129072.1), a copy-number

inducible derivative of plasmid R1. The vector was ligated

with adaptors complementary to the insert adaptors and

purified. The sheared, adapted mouse DNA was annealed to

adapted vector DNA, and transformed into

chemically-competent E. coli XL10-Gold (Stratagene) cells

and selected for ampicillin resistance."

BASE COUNT 106 a 31 c 33 g 72 t
ORIGIN

Query Match 1.1%; Score 21; DB 17; Length 242;

Best Local Similarity 100.0%; Pred. No. 41;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 132 TGAATAATCATGCCAGTGTCT 152

|||||

Db 44 TGAATAATCATGCCAGTGTCT 64

RESULT 5
BB283826/c

LOCUS

DEFINITION BB283826 RIKEN full-length enriched, adult retina Mus musculus cDNA

clone A30103H22 3', mRNA sequence.

ACCESSION BB283826

VERSION BB283826.1 GI:8984275

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 533)

Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1473 ATACTTGATCATTTTGAT 1492
|||||
Db 414 ATACTTGATCATTTTGAT 433

RESULT 10
LOCUS A0338598 449 bp DNA linear GSS 12-JUN-1999
DEFINITION HS 2202 B1 C01 MR CIT Approved Human Genomic Sperm Library D Homo
Sapiens genomic clone Plate=2202 Col=1 Row=F, DNA sequence.
ACCESSION A0338598
VERSION A0338598.1 GI:4144614
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 449)
Mahairas,G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
99380589
Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones may be purchased from Research Genetics (info@resgen.com).
BAC end Web Server: http://www.htsc.washington.edu
Plate: 2202 row: F column: 1
Seq primer: M13 Reverse
Class: BAC ends
High quality sequence stop: 449.
Location/Qualifiers
1..449

FEATURES
source
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_plate="2202 Col=1 Row=F"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBelOBAC11, BAC Clones in
E-Coli DH10B"

BASE COUNT 133 a 90 c 86 g 139 t 1 others
ORIGIN

Query Match 1.1%; Score 20; DB 17; Length 449;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1410 AATTCGAATGATTTGCA 1429
|||||
Db 171 AATTCGAATGATTTGCA 190

RESULT 11
LOCUS BI210777 497 bp mRNA linear EST 11-JUN-2001
DEFINITION EST528817 CTOS Lycopersicon esculentum cDNA clone CTOS24E24 5' end,
mRNA sequence.
ACCESSION BI210777
VERSION BI210777.1 GI:14688501
KEYWORDS EST.
SOURCE tomato.
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asterales; eunasterids II; Solanales; Solanaceae; Solanum;
Lycopersicon.
1 (bases 1 to 497)
van der Hoeven,R., Sun,H., Bezzerides,J., Cho,J., Uterback,R.,
Romling,C. and Tanksley,S.
Generation of ESTs from Tomato Suspension Cultures
Unpublished (2001)
Contact: CUGI
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: http://www.genome.clemson.edu/orders/index.html.
Location/Qualifiers
1..497

FEATURES
source
/organism="Lycopersicon esculentum"
/cultivar="TA496, E6203"
/db_xref="taxon:4081"
/clone_plate="CTOS24E24"
/clone_lib="CTOS"
/issue_type="suspension cultures"
/lab_host="SOLR"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI; Suspension cultures of L.esculentum E6203 were grown
in Murashige and Skoog based medium, supplemented with 15%
coconut milk (filter sterilized and added after
autoclaving), 2% saccharose, and 1mg/ml 2,4D (pH5.8).
Fresh medium was added every 7 days, and cultures were
grown at 25 C, with 12hrs of light and continuous
shaking."

BASE COUNT 139 a 84 c 109 g 165 t
ORIGIN

Query Match 1.1%; Score 20; DB 13; Length 497;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 835 TGGAGGTGTTTCTTTAG 854
|||||
Db 371 TGGAGGTGTTTCTTTAG 390

RESULT 12
LOCUS DR12E9S/c 507 bp DNA linear GSS 06-JUN-2002
DEFINITION Danio rerio genomic clone DKey-12E9, genomic survey sequence.
ACCESSION AL744698
VERSION AL744698.1 GI:21338981
KEYWORDS GSS.
SOURCE zebrafish.
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danio.
1 (bases 1 to 507)
Humphray,S.J., Huckle,E. and Hunt,S.E.
Direct Submission
Submitted (06-JUN-2002) The Sanger Institute, Wellcome Trust Genome
Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail contact:
humquerry@sanger.ac.uk Unpublished
This sequence was generated from the SP6 end of BAC 12E9. 12E9 is
part of the Daniokey Pilot BAC library created by R. Plasterk and
N.V. Keygene.
Further details: http://www.sanger.ac.uk/Projects/D_rerio/.

COMMENT

FEATURES
source
Location/Qualifiers
1..507
/organism="Danio rerio"
/db_xref="taxon:7955"
/clone_plate="DKey-12E9"
/issue_type="Testis"
/note="vector pindigobAC-536"

BASE COUNT 184 a 76 c 82 g 165 t
ORIGIN

```

Query Match          1.1%; Score 20; DB 17; Length 507;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1644 TCTTTAAATTTGCATTATT 1663
Db 338 TCTTTAAATTTGCATTATT 319

RESULT 13
LOCUS BH017124 658 bp DNA linear GSS 18-MAY-2001
DEFINITION TDGCT72TH CTG Lycopersicon esculentum genomic clone CTG18L24, DNA
sequence.
ACCESSION BH017124
VERSION BH017124.1 GI:14148162
KEYWORDS GSS,
SOURCE tomato.
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
Lycopersicon.
REFERENCE 1 (bases 1 to 658)
AUTHORS van der Hoeven, R., Sun, H., Cho, J., Utterback, T., Ronning, C. and
Tanksley, S.
TITLE Tomato Demethylated Genomic DNA Sequences
JOURNAL Unpublished (2001)
COMMENT Contact: CUGI
Clemson University Genomics Institute
100 Jordan Hall, Clemson, SC 29634, USA
Email: http://www.genome.clemson.edu/orders/index.html
tomato demethylated genomic DNA
Seq primer: M13P-R
Class: shotgun.
FEATURES             Location/Qualifiers
     source           1..658
     /organism="Lycopersicon esculentum"
     /cultivar="E6203"
     /db_xref="taxon:4081"
     /clone="CTG18L24"
     /clone_lib="CTG"
     /tissue_type="young leaves"
     /dev_stage="12-14 weeks post harvest"
     /lab_host="E.coli JM109"
     /note="Vector: pBluescript SK(-); Site 1: EcoRI; Site 2:
XhoI; This library was made from short EcoRI digested
fragments of the genome of Lycopersicon esculentum ligated
into pBS (SK-). The fragments were cloned into the
methylation restrictive E.coli strain JM109 with the
purpose of enriching the library for non-methylated DNA
fragments. This procedure may enrich the pool of cloned
fragments in JM109 cells for sequences representing
expressed genes. Average insert size 1.27 Kb."
BASE COUNT 159 a 141 c 143 g 215 t
ORIGIN

Query Match          1.1%; Score 20; DB 17; Length 658;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1199 AACGCCAAATTTTGTCAAA 1218
Db 517 AACGCCAAATTTTGTCAAA 536

RESULT 14
LOCUS CNS01NJ2/c 690 bp DNA linear GSS 14-JUN-2001
DEFINITION Anopheles gambiae GSS T7 end of clone 24D03 of NotreDamel library
from strain PEST of Anopheles gambiae (African malaria mosquito),

```

```

genomic survey sequence.
AL152287 830 bp DNA linear GSS 06-JUN-2001
VERSION AL152287.1 GI:7013206
KEYWORDS GSS,
SOURCE African malaria mosquito.
ORGANISM Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
Anopheles.
REFERENCE 1 (bases 1 to 690)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (16-FEB-2000) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : segret@genoscope.cns.fr)
- Web : www.genoscope.cns.fr)
REFERENCE 2 (bases 1 to 690)
AUTHORS Roth, C.W., Brey, P.T., Ke, Z., Collins, F.H. and Weissenbach, J.
TITLE Direct Submission
JOURNAL Submitted (16-FEB-2000) BBMI, Institut Pasteur, 25, rue du Dr.
Roux, Paris 75015, France
COMMENT This clone is from an A. gambiae BAC library provided by F.H.
Collins and sequenced by Genoscope in collaboration with the
Laboratory of Biochem. and Biol. Molec. of Insects, Institut
Pasteur.
FEATURES             Location/Qualifiers
     source           1..690
     /organism="Anopheles gambiae"
     /strain="PEST"
     /db_xref="taxon:7165"
     /clone="24D03"
     /clone_lib="NotreDamel"
     /notes="end : T7"
BASE COUNT 245 a 105 c 105 g 210 t 25 others
ORIGIN

Query Match          1.1%; Score 20; DB 17; Length 690;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1385 TTTATAGCACATTTCTTTT 1404
Db 658 TTTATAGCACATTTCTTTT 639

RESULT 15
LOCUS CNS01HVA/c 830 bp DNA linear GSS 06-JUN-2001
DEFINITION Anopheles gambiae GSS SP6 end of clone 09J12 of NotreDamel library
from strain PEST of Anopheles gambiae (African malaria mosquito),
genomic survey sequence.
ACCESSION AL144935
VERSION AL144935.1 GI:7003097
KEYWORDS GSS,
SOURCE African malaria mosquito.
ORGANISM Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
Anopheles.
REFERENCE 1 (bases 1 to 830)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (16-FEB-2000) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : segret@genoscope.cns.fr)
- Web : www.genoscope.cns.fr)
REFERENCE 2 (bases 1 to 830)
AUTHORS Roth, C.W., Brey, P.T., Ke, Z., Collins, F.H. and Weissenbach, J.
TITLE Direct Submission
JOURNAL Submitted (16-FEB-2000) BBMI, Institut Pasteur, 25, rue du Dr.
Roux, Paris 75015, France
COMMENT This clone is from an A. gambiae BAC library provided by F.H.
Collins and sequenced by Genoscope in collaboration with the
Laboratory of Biochem. and Biol. Molec. of Insects, Institut
Pasteur.

```

LOCUS	CNS004ZF	918 bp	DNA	linear	GSS 26-JUL-2000
DEFINITION	Tetradodon nigroviridis genome survey sequence SP6 end of clone 016P16 of library B from Tetradodon nigroviridis, genomic survey sequence.				
ACCESSION	ALJ12324				
KEYWORDS	GI:9545192				
SOURCE	GSS; genome survey sequence. Tetradodon nigroviridis Tetradodon nigroviridis Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Perciformes; Tetraodontiformes; Tetraodontidae; Tetradodon.				
REFERENCE	1 (bases 1 to 918) Roest Crolius,H., Jalllon,O., Dasilva,C., Bouneau,L., Fisher,C., Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F., Saurin,W., and Weissenbach,J. Estimate of human gene number provided by genome-wide analysis using Tetradodon nigroviridis DNA sequence Nat. Genet. 25 (2), 235-238 (2000)				
TITLE	JOURNAL MEDLINE PUBMED				
REFERENCE	20296633 10835645				
AUTHORS	2 (bases 1 to 918) Crolius,H.R., Jalllon,O., Dasilva,C., Ozouf-Costez,C., Fizames,C., Fischer,C., Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and Weissenbach,J. Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetradodon nigroviridis Genome Res. 10 (7), 939-949 (2000)				
TITLE	JOURNAL MEDLINE PUBMED				
REFERENCE	20359837 10899143				
AUTHORS	3 (bases 1 to 918) Genoscope. Direct Submission Submitted (12-APR-2000) This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetradodon nigroviridis genome. For more information, please take a look at http://www.genoscope.cns.fr/tetradodon. Location/Qualifiers				
FEATURES	source				
	1..918 /organism="Tetradodon nigroviridis" /db_xref="taxon:99883" /cclone="016P16" /cclone_1lb="B" /note="Genoscope sequence ID : COAB016DH08B1-end : SP6"				
BASE COUNT	192 a 208 c 205 g 277 t 36 others				
ORIGIN					
Query Match	1.1%; Score 20; DB 17; Length 918;				
Best Local Similarity	100.0%; Pred.No. 1.5e+02;				
Matches 20; Conservative	0; Mismatches 0; Indels 0; Gaps 0;				
OY	1055 ATCAAAATCTATTGTAGTGA 1074 				
Db	754 ATCAAATCTATTGTAGTGA 735				
RESULT 18	BF876881/c	169 bp	mRNA	linear	EST 17-JAN-2001
LOCUS	BF876881				
DEFINITION	PW4-EF0154-151100-004-a12 EF0154 Homo sapiens cDNA, mRNA sequence.				
ACCESSION	BF876881				
VERSION	BF876881.1				
KEYWORDS	GI:12267011				
SOURCE	EST. human.				
ORGANISM	Homo sapiens				
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.					
1 (bases 1 to 169)					
Nagai,M.A., Garcia Correa,R., Verjovski-Almeida,S., Bionnes,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bata,G.S., Simpson,D.H.,					
REFERENCE					
AUTHORS					

Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.
 Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
 Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
 20202663
 Contact: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
 (http://www.ludwig.org.br/scripts/gethtml2.pl?l1=PM4&l2=PM4-ET0154-151100-004-a12&l3=2000-11-15&l4=1)
 Seq primer: puc 18 forward
 High quality sequence stop: 169.

FEATURES

```

source
1. .169
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="ET0154"
/dev_stage="Adult"
/note="Organ: lung tumor; Vector: puc18; Site 1: SmaI; Site 2: SmaI; A mini-library was made by cloning products derived from ORSTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
39 c 43 g 32 t
55 a

```

BASE COUNT

```

ORIGIN
Query Match 1.0%; Score 19; DB 12; Length 169;
Best Local Similarity 100.0%; Pred. No. 4e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Qy 1280 TTGGCAATCATCTGCTCT 1298

Db 47 TTGGCAATCATCTGCTCT 29

RESULT 19

```

AV223805 205 bp mRNA linear EST 14-NOV-2001
LOCUS AV223805 RIKEN full-length enriched, 18 days pregnant, placenta and extra embryonic tissue Mus musculus cDNA clone 383041A10 3' similar to AF032668 Rattus norvegicus rsec15 mRNA, mRNA sequence.
AV223805
AV223805.1 GI:6173183

```

ACCESSION

VERSION

KEYWORDS

SOURCE

```

house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
1 (bases 1 to 205)

```

REFERENCE

AUTHORS

Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Carninci,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N., Hirozane,T., Hori,F., Ishii,Y., Ishikawa,T., Itoh,M., Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N., Kojima,Y., Koya,S., Kusakabe,M., Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y., Owa,C., Ozawa,Y., Saito,H., Sano,M., Sato,K., Shibata,K., Shibata,Y., Shigemoto,Y., Shiraki,T., Sogabe,Y., Sugahara,Y., Suzuki,H., Suzuki,H., Takahashi,F., Tateo,M., Tominega,N., Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yasunishi,A., Yokota,T., Yoshiki,A., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
 RIKEN Mouse ESTs (Konno,H., et al. 1999)
 Unpublished (1999)
 Contact: Yoshihide Hayashizaki

TITLE

COMMENT

Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 Tel: 81-45-503-9222
 Fax: 81-45-503-9216
 Email: genome-res@gscc.riken.go.jp,
 URL:http://genome.gsc.riken.go.jp/
 Sasaki,N., Izawa,M., Watahiki,M., Ozawa,K., Tanaka,T., Yoneda,Y., Matsuura,S., Carninci,P., Muramatsu,M., Okazaki,Y. and Hayashizaki,Y.
 Transcriptional sequencing: A method for DNA sequencing using RNA polymerase. Proc. Natl. Acad. Sci. U.S.A. 95 (7), 3455-3460 (1998)
 Itoh,M., Kitsuai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J., Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki,Y. and Hayashizaki,Y.
 Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999)
 Carninci,P. and Hayashizaki,Y.
 High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)
 Please visit our web site (http://genome.gsc.riken.go.jp) for further details.

FEATURES

```

source
1. .205
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone_lib="RIKEN full-length enriched, 18 days pregnant, placenta and extra embryonic tissue"
/sex="female"
/tissue_type="placenta and extra embryonic tissue"
/dev_stage="18 days pregnant, adult"
/lab_host="DH108"
/note="Site 1: Sall; Site 2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAGATCCAGAGCTCTTTTTTTTTTTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 50.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCGAGTTAATAATTAATCCCCCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I"
77 a 29 c 24 g 75 t

```

BASE COUNT

ORIGIN

```

Query Match 1.0%; Score 19; DB 9; Length 205;
Best Local Similarity 100.0%; Pred. No. 4.1e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Qy 1400 TTTTGAAGAAAATTCAAA 1418

Db 133 TTTTGAAGAAAATTCAAA 151

RESULT 20

```

BH342232/c
LOCUS BH342232 236 bp DNA linear GSS 03-DEC-2001
DEFINITION CH230-9J13 TJ CHORI-230 Segment 1 Rattus norvegicus genomic clone CH230-9J13, DNA sequence.
ACCESSION BH342232
VERSION BH342232.1 GI:17272966
KEYWORDS GSS.
SOURCE Norway rat.

```

ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE 1 (bases 1 to 236)
Zhao,S., Shetty,J., Shatman,S., Tesgaye,G., Geer,K., Shvartsbeyn
A., Gebregorgis,E., Overton,L., Russell,D., Chen,D., Riggs,F., de
Jong,P. and Fraser,C.W.
Rat BAC End Sequences from Library CHORI-230 EcoRI segment
Unpublished (1999)
Other GSSs: CH230-9J13.TV

TITLE JOURNAL
COMMENT

FEATURES
source
Location/Qualifiers
1..236
/organism="Rattus norvegicus"
/strain="BN/SnHsd/MCw"
/db_xref="taxon:10116"
/clone="CH230-9J13"
/clone_1b="CHORI-230 Segment 1"
/sex="Female"
/cell_type="Brain"
/note="Vector: PTARBA2.1; Site 1: EcoRI; Site 2: EcoRI;
CHORI-230 Rat (BN/SnHsd/MCw) BAC library produced by
Pleier de Jong"
Base Count 68 a 55 c 66 g 47 t

Query Match 1.0%; Score 19; DB 17; Length 236;
Best Local Similarity 100.0%; Pred. No. 4.2e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 635 GATACAGGCGCTATTAAAG 653
|||||
Db 93 GATACAGGCGCTATTAAAG 75

RESULT 21
BE527201 246 bp mRNA linear EST 19-MAR-2001
LOCUS M67F07STM Arabidopsis developing seed Arabidopsis thaliana cDNA
DEFINITION
Accession BE527201
Version BE527201.1 GI:9785179
KEYWORDS
SOURCE EST.
ORGANISM thale cress
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
1 (bases 1 to 246)
White,J.A., Todd,J., Newman,T., Rocks,N., Gitzke,T., Martinez de
Iliarduya,O., Jaworski,J.G., Ohlrogge,U. and Benning,C.
A new set of Arabidopsis expressed sequence tags from developing
seeds. The metabolic pathway from carbohydrates to seed oil
Plant Physiol. 124 (4), 1582-1594 (2000)
Contact: Benning, C
Dept. of Biochemistry & Molecular Biology

Michigan State University
224 Biochemistry, Michigan State University, East Lansing, MI 48824
, USA
Tel: 517 355 1609
Fax: 517 353 9334
Email: benning@msu.edu
Clones were originally prepared at Michigan State University.
Arabidopsis Biological Resource Center, The Ohio State University,
309 Botany & Zoology Bldg., 1735 Neil Avenue, Columbus, OH 43210
USA. FAX: 6142920603 TEL: 6142929371.

FEATURES
source
Location/Qualifiers
1..246
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="600036027R1"
/clone_1b="Arabidopsis developing seed"
/tissue_type="seed"
/dev stage="5-13 days after flowering"
/lab_host="E.coli"
/note="Organ: Developing seed; Vector: pBluescript SK-;
Site_1: EcoRI; Site_2: XhoII"
Base Count 66 a 63 c 54 g 63 t

Query Match 1.0%; Score 19; DB 10; Length 246;
Best Local Similarity 100.0%; Pred. No. 4.2e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1478 TTGATCATTTTGTATCGTA 1496
|||||
Db 74 TTGATCATTTTGTATCGTA 92

RESULT 22
A1445399 262 bp mRNA linear EST 13-APR-1999
LOCUS t1j24h08.x1 NCI CGAP Gas4 Homo sapiens cDNA clone IMAGE:2142495 3'
DEFINITION similar to gb:M24095 HLA CLASS I HISTOCOMPATIBILITY ANTIGEN,
A-26 (A-10) A*2601 ALPHA (HUMAN);, mRNA sequence.
Accession A1445399
Version A1445399.1 GI:4287940
KEYWORDS
SOURCE EST.
ORGANISM human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 262)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgaabs-r@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
www.bio.lnl.gov/bbrp/image/image.html
Insert Length: 962 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 1.
Location/Qualifiers
1..262
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2142495"
/clone_1b="NCI CGAP Gas4"
/tissue_type="poorly differentiated adenocarcinoma with
signet ring cell features"

```

/lab_host="DH10B"
/Note="Organ: stomach; Vector: pCMV-SPORT6; Site 1: SalI;
Site 2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.69 kb. Life Technologies catalog #:
11549-011"

BASE COUNT      81 a   47 C   63 G   71 T
ORIGIN

Query Match      1.0%; Score 19; DB 9; Length 262;
Best Local Similarity 100.0%; Pred. No. 4.2e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy  671 ATTTAGGAATTTCTCTT 689
Db   19 ATTTAGGAATTTCTCTT 37

RESULT 23
LOCUS      AV155678      286 bp      mRNA      linear      EST 08-JUL-1999
DEFINITION AV155678 Mus musculus hippocampus C57BL/6J adult Mus musculus cDNA
clone 2900083G01, mRNA sequence.
ACCESSION  AV155678
VERSION     AV155678.1 GI:5427095
KEYWORDS    EST.
SOURCE      house mouse.
ORGANISM    Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
1 (bases 1 to 286)
Akhira,S., Akiyama,J., Fukuda,S., Fukunishi,Y., Funayama,T., Hara
,A., Hayatsu,N., Hori,F., Ishikawa,T., Itoh,M., Izawa,M., Kawai,J.,
Kikuchi,N., Kojima,Y., Matsuyama,T., Nitsuma,H., Oda,H., Owa,C.,
Sato,K., Shibata,Y., Shigemoto,Y., Shiraki,T., Sogabe,Y., Sugahara
,Y., Suzuki,H., Suzuki,H., Tateno,M., Tomaru,Y., Tominaga,N.,
Watanabe,S., Yagame,M., Yamamura,T., Yokota,T., Yoshino,M.,
Muramatsu,M., Okazaki,Y. and Hayashizaki,Y.
RIKEN Mouse ESTs
Unpublished (1999)
Contact: Chie Owa
Genome Science Laboratory
RIKEN
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: 81-298-36-9145
Fax: 81-298-36-9098
Email: genome-res@rtc.riken.go.jp
Thermostabilization and thermoactivation of thermostable enzymes by
trehalose and its application for the synthesis of full length cDNA
(Proc. Natl. Acad. Sci. U.S.A. 95(2):520-524 (1998))
Transcriptional sequencing: A method for DNA sequencing using RNA
polymerase (Proc. Natl. Acad. Sci. U.S.A. 95(7):3455-3460 (1998))
Please visit our web site (http://genome.rtc.riken.go.jp) for
further details.
Location/Qualifiers
1. .286
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="2900083G01"
/clone_lib="Mus musculus hippocampus C57BL/6J adult"
/sex="male"
/tissue_type="hippocampus"
/dev_stage="adult"
/Note="Organ: mammary gland; Vector: pT73D-Pac (Pharmacia
) with a modified polylinker; Site 1: Not I; Site 2: Eco
RI; 1st strand cDNA was primed with a Not I - oligo(dT)
primer [5',
TGTACCAATCTGAAGTGGGCGCGCGAATGTTTTTTTTTTTTTTTTTTTTT
T 3']; double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pT73 vector.
RNA provided by Dr. Minoru Ko, Wayne State Univ. Library

/lab_host="DH10B"
/Note="Organ: stomach; Vector: pCMV-SPORT6; Site 1: SalI;
Site 2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.69 kb. Life Technologies catalog #:
11549-011"

BASE COUNT      81 a   47 C   63 G   71 T
ORIGIN

Query Match      1.0%; Score 19; DB 9; Length 286;
Best Local Similarity 100.0%; Pred. No. 4.3e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy  894 TGGGAATGGATTTTAAAGG 912
Db   150 TGGGAATGGATTTTAAAGG 168

RESULT 24
LOCUS      AV138833      298 bp      mRNA      linear      EST 01-JUL-1999
DEFINITION AV138833 Mus musculus C57BL/6J 10-11 day embryo Mus musculus cDNA
clone 2810046P20, mRNA sequence.
ACCESSION  AV138833
VERSION     AV138833.1 GI:5324563
KEYWORDS    EST.
SOURCE      house mouse.
ORGANISM    Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 298)
Carninci,P., Shibata,K., Ozawa,Y., Konno,H., Itoh,M., Aizawa,K.,
Akhira,S., Akiyama,J., Fukuda,S., Fukunishi,Y., Funayama,T., Hara
,A., Hayatsu,N., Hori,F., Ishikawa,T., Itoh,M., Izawa,M., Kawai,J.,
Kikuchi,N., Kojima,Y., Matsuyama,T., Nitsuma,H., Oda,H., Owa,C.,
Sato,K., Shibata,Y., Shigemoto,Y., Shiraki,T., Sogabe,Y., Sugahara
,Y., Suzuki,H., Suzuki,H., Tateno,M., Tomaru,Y., Tominaga,N.,
Watanabe,S., Yagame,M., Yamamura,T., Yokota,T., Yoshino,M.,
Muramatsu,M., Okazaki,Y. and Hayashizaki,Y.
RIKEN Mouse ESTs
Unpublished (1999)
Contact: Chie Owa
Genome Science Laboratory
RIKEN
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: 81-298-36-9145
Fax: 81-298-36-9098
Email: genome-res@rtc.riken.go.jp
Thermostabilization and thermoactivation of thermostable enzymes by
trehalose and its application for the synthesis of full length cDNA
(Proc. Natl. Acad. Sci. U.S.A. 95(2):520-524 (1998))
Transcriptional sequencing: A method for DNA sequencing using RNA
polymerase (Proc. Natl. Acad. Sci. U.S.A. 95(7):3455-3460 (1998))
Please visit our web site (http://genome.rtc.riken.go.jp) for
further details.
Location/Qualifiers
1. .298
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="2810046P20"
/clone_lib="Mus musculus C57BL/6J 10-11 day embryo"
/sex="mixed"
/dev_stage="10-11 day embryo"
/Note="Organ: mammary gland; Vector: pT73D-Pac (Pharmacia
) with a modified polylinker; Site 1: Not I; Site 2: Eco
RI; 1st strand cDNA was primed with a Not I - oligo(dT)
primer [5',
TGTACCAATCTGAAGTGGGCGCGCGAATGTTTTTTTTTTTTTTTTTTTTT
T 3']; double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pT73 vector.
RNA provided by Dr. Minoru Ko, Wayne State Univ. Library
constructed and normalized by Bento Soares and M.Fatima
Bonaldo."

```

BASE COUNT
ORIGIN

92 a 47 c 63 g 96 t

Query Match 1.0%; Score 19; DB 9; Length 298;
 Best Local Similarity 100.0%; Pred. No. 4.3e+02;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 894 TGGGATGATTTTAAAGG 912
 |||||
 Db 214 TGGGATGATTTTAAAGG 232

RESULT 25
 A1556340/C 299 bp mRNA linear EST 23-MAR-1999

LOCUS A1556340
 DEFINITION UI-R-C2P-r9-c-07-0-UI.s1 UI-R-C2P Rattus norvegicus cDNA clone
 UI-R-C2P-r9-c-07-0-UI 3', mRNA sequence.

ACCESSION A1556340
 A1556340.1 GI:4488703

VERSION
 KEYWORDS
 SOURCE Norway rat.
 ORGANISM Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 1 (bases 1 to 299)
 Bonaldo,M.F., Lennon,G. and Soares,M.B.

REFERENCE
 AUTHORS
 TITLE Normalization and subtraction: two approaches to facilitate gene
 discovery
 Genome Res. 6 (9), 791-806 (1996)

JOURNAL
 MEDLINE
 COMMENT Contact: Soares, MB
 Program for Rat Gene Discovery and Mapping
 University of Iowa
 451 Eckstein Medical Research Building Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565

Email: meares@blue.weeg.uiowa.edu
 The sequence tag present in the cDNA between the NotI site and the
 oligo-dt track served to identify it as a clone from the normalized
 adult muscle library. cDNA library preparation: M. Fatima Bonaldo,
 Ph.D. Clone distribution: clones will be available through Research
 Genetics (www.resgen.com).
 Seq primer: M13 Forward.

FEATURES
 source Location/Qualifiers

1. 299
 /organism="Rattus norvegicus"
 /strain="Sprague-Dawley"
 /db_xref="taxon:10116"
 /clone="UI-R-C2P-r9-c-07-0-UI"
 /clone_1lb="UI-R-C2P"
 /dev_stage="adult"
 /lab_host="DH10B (Life Technologies)"
 /note="Vector: pT73D-Pac (Pharmacia) with a modified
 polylinker; Site 1: Not I; Site 2: Eco RI; The UI-R-C2P
 library is a subtracted library derived from the UI-R-C1
 library, which is a subtracted library derived from the
 UI-R-C0 library. The UI-R-C0 library consisted of a
 mixture of individually tagged normalized libraries
 constructed from rat placenta, adult lung, brain, liver,
 kidney, heart, spleen, ovary, muscle, 8, 12 and 18-day
 embryo. The tag is a string of 3-5 nucleotides present
 between the Not I site and the oligo-dt track which allows
 identification of the library of origin of a clone within
 the mixture. The subtracted library (UI-R-C2P) was
 constructed as follows: PCR amplified cDNA inserts from
 UI-R-C1 clones from which 3' ESTs had been derived was
 used as a driver in a hybridization with the UI-R-C1
 library in the form of single-stranded circles. The
 remaining single-stranded circles (subtracted library) was
 purified by hydroxyapatite column chromatography,
 converted to double-stranded circles and electroporated
 into DH10B bacteria (Life Technologies) to generate the
 UI-R-C2P library. This procedure has been previously
 described (Bonaldo, Lennon and Soares, Genome Research 6:

BASE COUNT 791-806, 1996)"
 ORIGIN
 93 a 41 c 45 g 120 t

Query Match 1.0%; Score 19; DB 9; Length 299;
 Best Local Similarity 100.0%; Pred. No. 4.3e+02;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1842 TTTATGACAAATTCAGATA 1860
 |||||
 Db 138 TTTATGACAAATTCAGATA 120

RESULT 26
 A146527
 LOCUS A146527

DEFINITION AV146527 Mus musculus C57BL/6J 10-11 day embryo Mus musculus cDNA
 clone 2810454K18, mRNA sequence.

ACCESSION AV146527.1 GI:5350660

VERSION
 KEYWORDS
 SOURCE house mouse.
 ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 306)
 Carninci,P., Shibata,K., Ozawa,Y., Konno,H., Itoh,M., Aizawa,K.,
 Akahira,S., Akiyama,J., Fukuda,S., Fukunishi,Y., Funayama,T., Hara
 A., Hayatsu,N., Hori,F., Ishikawa,T., Itoh,M., Izawa,M., Kawai,J.,
 Kikuchi,N., Kojima,Y., Matsuyama,T., Nitsuma,H., Oda,H., Owa,C.,
 Sato,K., Shibata,Y., Shigemoto,Y., Shiraki,T., Sogabe,Y., Sugahara
 Y., Suzuki,H., Suzuki,H., Tateo,M., Tomaru,Y., Tomimaga,N.,
 Watanabe,S., Yagame,M., Yamamura,T., Yokota,T., Yoshino,M.,
 Muramatsu,M., Okazaki,Y. and Hayashizaki,Y.

REFERENCE
 AUTHORS
 TITLE RIKEN Mouse ESTs
 JOURNAL Unpublished (1999)
 COMMENT Contact: Chie Owa
 Genome Science Laboratory
 RIKEN
 3-1-1 Koyada, Tsukuba, Ibaraki 305-0074, Japan
 Tel: 81-298-36-9145
 Fax: 81-298-36-9098
 Email: genome-res@rtc.riken.go.jp

Thermostabilization and thermocyclization of thermostable enzymes by
 trehalose and its application for the synthesis of full length cDNA
 (Proc. Natl. Acad. Sci. U.S.A. 95(2):520-524 (1998))
 Transcriptional sequencing: A method for DNA sequencing using RNA
 polymerase (Proc. Natl. Acad. Sci. U.S.A. 95(7):3455-3460 (1998))
 please visit our web site (http://genome.rtc.riken.go.jp) for
 further details.

FEATURES
 source Location/Qualifiers

1. 306
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="2810454K18"
 /clone_1lb="Mus musculus C57BL/6J 10-11 day embryo"
 /sex="mixed"
 /dev_stage="10-11 day embryo"
 /note="Organ: mammary gland; Vector: pT73D-Pac (Pharmacia
) with a modified polylinker; Site 1: Not I; Site 2: Eco
 RI; 1st strand cDNA was primed with a Not I - oligo(dT)
 primer [5']
 TTTTCCATCGAAGTGGAGCGCGCGAATGTTTTTTTTTTTTTTTTTTTTT
 T 3'; double-stranded cDNA was ligated to Eco RI
 adaptors (Pharmacia), digested with Not I and cloned into
 the Not I and Eco RI sites of the modified pT73 vector.
 RNA provided by Dr. Minoru Ko, Wayne State Univ. Library
 constructed and normalized by Bento Soares and M. Fatima
 Bonaldo."

BASE COUNT 94 a 58 c 68 g 86 t
 ORIGIN

```

Query Match      1.0%; Score 19; DB 9; Length 306;
Best Local Similarity 100.0%; Pred. No. 4.3e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 894 TGGGAATGGATTTTAAAGG 912
|||||
Db 225 TGGGAATGGATTTTAAAGG 243

RESULT 27
AV117891
LOCUS
DEFINITION
AV117891 Mus musculus C57BL/6J 10-day embryo Mus musculus cDNA
clone 2610205N14, mRNA sequence.
ACCESSION
AV117891
VERSION
AV117891.1 GI:5300042
KEYWORDS
EST.
SOURCE
human mouse.
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 319)
AUTHORS
Carninci, P., Shibata, K., Ozawa, Y., Konno, H., Itoh, M., Aizawa, K.,
A., Hayatsu, N., Hori, F., Ishikawa, T., Itoh, M., Izawa, M., Kawai, J.,
Kikuchi, N., Kojima, Y., Matsuyama, T., Niitsuma, H., Oda, H., Owa, C.,
Sato, K., Shibata, Y., Shigemoto, Y., Shiraki, T., Sogabe, Y., Sugahara
Y., Suzuki, H., Suzuki, H., Tateno, M., Tomaru, Y., Tomimaga, N.,
Watanabe, S., Yagame, M., Yamamura, T., Yokota, T., Yoshino, M.,
Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.
TITLE
RIKEN Mouse ESTs
JOURNAL
Unpublished (1999)
COMMENT
Contact: Chie Owa
Genome Science Laboratory
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: 81-298-36-9145
Fax: 81-298-36-9098
Email: genome-res@rtc.riken.go.jp
Thermotabilization and thermoactivation of thermostable enzymes by
trehalose and its application for the synthesis of full length cDNA
(Proc. Natl. Acad. Sci. U.S.A. 95(2):520-524 (1998))
Transcriptional sequencing: A method for DNA sequencing using RNA
polymerase (Proc. Natl. Acad. Sci. U.S.A. 95(7):3455-3460 (1998))
Please visit our web site (http://genome.rtc.riken.go.jp) for
further details.
FEATURES
source
Location/Qualifiers
1..319
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="2610205N14"
/clone_lib="Mus musculus C57BL/6J 10-day embryo"
/sex="mixed"
/dev_stages="10-day embryo"
/notes="Organ: mammary gland; Vector: pT7T3D-Pac (Pharmacia
RI) with a modified polylinker; Site 1: Not I; Site 2: Eco
RI; 1st strand cDNA was primed with a Not I - oligo (dT)
primer [5',
TGTACCACTGCTGAAGTGGGCGCGCGCGGATCGTTTTTTTTTTTTTTTTTTTTTTT
T 3']; double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pT7T3 vector.
RNA provided by Dr. Minoru Ko, Wayne State Univ. Library
constructed and normalized by Bento Soares and M. Fatima
Bonaldo."
BASE COUNT 93 a 58 c 65 g 103 t
ORIGIN
Query Match      1.0%; Score 19; DB 9; Length 319;
Best Local Similarity 100.0%; Pred. No. 4.3e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 894 TGGGAATGGATTTTAAAGG 912
|||||
Db 225 TGGGAATGGATTTTAAAGG 243

RESULT 28
AA081640
LOCUS
DEFINITION
AA081640 332 bp mRNA linear EST 21-OCT-1996
zn20c03.r1 Stratagene neuroepithelium NT2RAMI 937234 Homo sapiens
cDNA clone IMAGE:547972 5', mRNA sequence.
ACCESSION
AA081640
VERSION
AA081640.1 GI:1623698
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 332)
AUTHORS
Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chispe, B.,
Chispe, S., Dietrich, N., DuBuque, T., Favell, A., Gish, W., Hawkins
M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N., Mardis, E., Moore
B., Morris, M., Parsons, J., Prange, C., Rifkin, L., Rohlfing, T.,
Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J., Trevas, E.,
Underwood, K., Wohldmann, P., Waterston, R., Wilson, R. and Marra, M.
Generation and analysis of 280,000 human expressed sequence tags
Genome Res. 6 (9), 807-828 (1996)
97044478
COMMENT
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -28W13 rev2 from Amersham.
FEATURES
source
Location/Qualifiers
1..332
/organism="Homo sapiens"
/db_xref="GDB:3926488"
/db_xref="taxon:9606"
/clone="IMAGE:547972"
/clone_lib="Stratagene neuroepithelium NT2RAMI 937234"
/dev_stage="Ntera-2/RA-MI neuroepithelial cells"
/lab_host="SOLR (kanamycin resistant)"
/notes="Vector: pBluescript SK-; Site 1: EcoRI; Site 2:
XhoI; Cloned unidirectionally.
(Ntera-2/ci.D1) precursor cells induced with Retinoic
Acid for 1 week, followed by 3 weeks in mitotic inhibitors
(Replate #2). Average insert size: 1.1 kb; Uni-ZAP XR
Vector; -5' adaptor sequence: 5' GAATTCGGCAGAG 3' -3'
adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3'"
BASE COUNT 92 a 77 c 78 g 83 t 2 others
ORIGIN
Query Match      1.0%; Score 19; DB 9; Length 332;
Best Local Similarity 100.0%; Pred. No. 4.3e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 370 AGCTGTCTTCTGAATCTT 388
|||||
Db 236 AGCTGTCTTCTGAATCTT 254

RESULT 29
AI466079
LOCUS
DEFINITION
AI466079 350 bp mRNA linear EST 09-MAR-1999
vW39d07.Y1 Soares mammary gland NbWVG Mus musculus cDNA clone
IMAGE:1246189 5', mRNA sequence.
ACCESSION
AI466079
VERSION
AI466079.1 GI:4320109
KEYWORDS
EST.
SOURCE
house mouse.

```

ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Scuriognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 350)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
Tumor Gene Index
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Unpublished (1997)
CONTACT: Robert Strausberg, Ph.D.
Email: cgapsb-r@mail.nih.gov
This clone is available royalty-free through INL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:659877
This read is a RESEQUENCE of a previously sequenced mouse clone
This read has been verified (found to hit its original self in the correct orientation)
Seq primer: -40RP from Gibco
High quality sequence stop: 333.
Location/Qualifiers

FEATURES
source
1..350
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone_1lb="IMAGE:1246189"
/clone_1lb="Soares_mammary_gland_NbMNG"
/sex="male"
/tissue_type="mammary_gland"
/dev_stage="4 weeks"
/lab_host="DH10B"
/note="Organ: mammary gland; Vector: pRTT3D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - 01igo (dT) primer [5']
TGTACCAATCTGAGTGGAGCGCGCGAGATGTTTTTTTTTTTTTTTTTT
T 3']; double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pRTT3 vector.
RNA provided by Dr. Minoru Ko, Wayne State Univ. Library
constructed and normalized by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 83 a 68 c 86 g 112 t 1 others
ORIGIN

Query Match 1.0%; Score 19; DB 9; Length 350;
Best Local Similarity 100.0%; Pred. No. 4.4e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1645 CTTTAATTTTGCATTATT 1663
Db 236 CTTTAATTTTGCATTATT 254

RESULT 30
AM436624/c 353 bp mRNA linear EST 09-JUL-2000
LOCUS AM436624
DEFINITION 77002 MARC 2P1G Sus scrofa cDNA 5', mRNA sequence.
ACCESSION AM436624
VERSION AM436624.1 GI:6971930
KEYWORDS EST.
SOURCE pig.
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.
REFERENCE 1 (bases 1 to 353)
Fahrenkrug, S.C., Fekking, B.A., Rohrer, G.A., Smith, T.P.L., Casas, E., Stone, R.T., Heaton, M.P., Grose, W.M., Bennett, G.A., Laegreid, W.M. and Keeler, J.W.
Design and use of two pooled tissue normalized cDNA libraries for EST discovery in swine
Unpublished (2000)
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA

Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@mail.marc.usda.gov
Single pass sequencing. Bases called and trimmed with phred v0.980904.e. Vector identified by cross_match with the -minscore 20 and -mismatch 12 options.
PCR Primers
FORWARD: AGGAACAGCTATGACCAT
BACKWARD: GTTTCCCACTACGACG
Plate: 32 row: N column: 5
Seq primer: ATTGATGACACTATAG.
Location/Qualifiers

FEATURES
source
1..353
/organism="Sus scrofa"
/db_xref="taxon:9823"
/clone_1lb="MARC 2P1G"
/tissue_type="pooled"
/lab_host="DH10B"
/note="Vector: PCMV SPORT6; Site 1: NotI; Site 2: SalI; Library made from pooled tissue from testis, ovary, endometrium, hypothalamus, pituitary, and placenta."

BASE COUNT 102 a 81 c 75 g 95 t
ORIGIN

Query Match 1.0%; Score 19; DB 10; Length 353;
Best Local Similarity 100.0%; Pred. No. 4.4e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 674 TTAGAAATTTCTCTTCAA 692
Db 248 TTAGAAATTTCTCTCAA 230

RESULT 31
BF364126 356 bp mRNA linear EST 24-NOV-2000
LOCUS BF364126
DEFINITION QV0-NN1022-230600-282-c09 NN1022 Homo sapiens cDNA, mRNA sequence.
ACCESSION BF364126
VERSION BF364126.1 GI:11326151
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 356)
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H., Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.
Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
CONTACT: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(<http://www.ludwig.org.br/scripts/gethtml2.pl?i=QV0&c2=QV0-NN1022-230600-282-c09&c3=2000-06-23&t=1>)
Seq primer: puc 18 forward
High quality sequence start: 7
High quality sequence stop: 27.
Location/Qualifiers
1..356
/organism="Homo sapiens"

FEATURES
source

```

/db_xref="taxon:9606"
/clone_lib="NN1022"
/dev_stage="Adult"
/note="Organ: nervous normal; Vector: puc18; Site 1: SmaI;
Site 2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 396,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
BASE COUNT      111 a      74 c      77 g      94 t
ORIGIN
Query Match      1.0%; Score 19; DB 12; Length 356;
Best Local Similarity 100.0%; Pred. No. 4.4e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 370 AGCTGTTCTTCTGAATCTT 388
Db 151 AGCTGTTCTTCTGAATCTT 169

RESULT 32
AZ537688/c
LOCUS      AZ537688      367 bp      DNA      linear      GSS 06-NOV-2000
DEFINITION      AST-DHW(7) Genetrapp PC-3 Human Prostatic Carcinoma Library Homo
sapiens genomic 5', DNA sequence.
ACCESSION      AZ537688
VERSION      AZ537688.1 GI:11115054
KEYWORDS      GSS.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 367)
AUTHORS      Henkel, G., Liyanage, M., Pratt, E., Huang, D., Riley, M., Bernardino, A.,
Durick, K. and Pollok, B.
TITLE      Exon-trap tags from a PC-3 GenomesScreen(TM) Library
JOURNAL      Unpublished (2000)
COMMENT      Contact: Greg Henkel
Gene Expression
Aurora Biosciences Corp.
11010 Torreyana Road, San Diego, CA 92121, USA
Tel: 8584048436
Fax: 8584046719
Email: henkelg@aurorabio.com
Pools of cells were isolated from a GenomesScreen(TM) library. The
library of cells was generated by retroviral integration of a gene
tagging element consisting of: 1) A promoterless beta-lactamase
preceded by a splice acceptor as a reporter for gene expression;
2) A promoter driving neomycin resistance followed by a splice
donor to trap downstream exons. 3' RACE from neomycin gene was
performed using total RNA from isolated pools. Output was shotgun
cloned in pAMP-1 and used to transform DH5-alpha competent
bacteria. 5' ends of reported sequences were immediately preceded
by splice donor from the trapping construct.
Class: exon-trapped.
FEATURES
source
1. 367
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Genetrapp PC-3 Human Prostatic Carcinoma
Library"
/tissue_type="Adenocarcinoma"
/cell_type="Epithelial"
/cell_line="PC-3"
/note="Organ: Prostate; Vector: pAMP-1; 3' RACE of total
RNA from genetrapp pools; shotgun clone in pAMP-1 and used
to transform DH5-alpha competent bacteria."
BASE COUNT      109 a      79 c      96 g      83 t
ORIGIN
Query Match      1.0%; Score 19; DB 17; Length 367;

```

```

Best Local Similarity 100.0%; Pred. No. 4.4e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 609 TGGCCTCAGCACTTCTTTT 627
Db 62 TGGCCTCAGCACTTCTTTT 44

RESULT 33
BQ528747
LOCUS      BQ528747      376 bp      mRNA      linear      EST 11-JUN-2002
DEFINITION      lab_Zea_mays cDNA, mRNA sequence.
ACCESSION      BQ528747
VERSION      BQ528747.1 GI:21390698
KEYWORDS      EST.
SOURCE      Zea mays.
ORGANISM      Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE      1 (bases 1 to 376)
AUTHORS      Walbot, V.
TITLE      Maize ESTs from various cDNA libraries sequenced at Stanford
University
JOURNAL      Unpublished (1999)
COMMENT      Contact: Walbot V
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 3524.1.37.1 row: H column: 11.
FEATURES
source
1. 376
/organism="Zea mays"
/db_xref="taxon:4577"
/clone_lib="3524 - Mature pollen from Sheila McCormick's
lab"
/tissue_type="pollen"
/dev_stage="mature"
/lab_host="SOLR"
/note="Vector: Stratagene's Uni-Zap XR (pBluescript SK-);
Site 1: EcoRI (5-prime); Site 2: XhoI (3-prime);
Unamplified cDNA library directionally cloned by Rima
Kulikauskas using Stratagene's Uni-Zap system. Insert
sizes ranged from 0.5Kb to 2Kb. 50 microliter aliquot had
338,000 pfu when it was made in Sept, 1995, from oligo
dT-primed poly A+ RNA."
BASE COUNT      100 a      92 c      97 g      87 t
ORIGIN
Query Match      1.0%; Score 19; DB 14; Length 376;
Best Local Similarity 100.0%; Pred. No. 4.4e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 374 GTTCTTCTGAATCTTTTGC 392
Db 269 GTTCTTCTGAATCTTTTGC 287

RESULT 34
BJ541842/c
LOCUS      BJ541842      389 bp      mRNA      linear      EST 09-AUG-2002
DEFINITION      BJ541842 MFO1SSB cDNA Oryzias latipes clone MFO1SSB036B15 3',
mRNA sequence.
ACCESSION      BJ541842
VERSION      BJ541842.1 GI:22200654
KEYWORDS      EST.
SOURCE      Japanese medaka.
ORGANISM      Oryzias latipes

```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha; Belontiiformes; Adrianchthyidae; Oryziinae; Oryzias.

REFERENCE 1 (bases 1 to 389)

Kohara,Y., Shin-i,T., Kimura,T., Narita,T., Jindo,T. and Takeda,H.

Medaka EST Project in Takeda's lab

Unpublished (2001)

Contact: Tadao Shin-i

Center For Genetic Resource Information

National Institute of Genetics

1111 Yata, Mishima,

Shizuoka 411-8540, Japan

Tel: 81-559-81-6856

Fax: 81-559-81-6855

Email: tshin@genes.nig.ac.jp.

Location/Qualifiers

1. .389

/organism="Oryzias latipes"

/db_xref="taxon:8090"

/clone_lib="MF01SSB036B15"

/sex="mixture of female and male"

/tissue_type="whole embryo"

/dev_stage="segmentation stage 20 - 25"

121 a 97 c 104 g 67 t

Query Match 1.0%; Score 19; DB 13; Length 389;

Best Local Similarity 100.0%; Pred. No. 4.4e+02;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 849 TTTTAGGTTGCGTTCCTCT 867

Db 143 TTTTAGGTTGCGTTCCTCT 125

RESULT 35

LOCUS T27099 396 bp mRNA linear EST 25-JUL-1996

DEFINITION NIBT285H12R Infant brain, LNL array of Dr. M. Soares INIB Homo

sapiens cDNA clone LAB285H12 5' end, mRNA sequence.

ACCSSION T27099

VERSION T27099.1 GI:774141

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 396)

Sikela,J.M.

Human cDNAs from infant brain

Unpublished (1993)

Contact: Sikela JM

Department of Pharmacology

University of Colorado Health Sciences Center

Box C236, 4200 E. 9th Ave, Denver CO 80262-0236

Tel: 3032708637

Fax: 3032707097

Email: mikki@tally.uchsc.edu

Seq primer: M13 Reverse.

Location/Qualifiers

1. .396

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_lib="LAB285H12"

/clone_lib="Infant brain, LNL array of Dr. M. Soares

INIB"

/note="Vector: Jafmid BA; Site 1: HindIII; Site 2: NotI;

Normalized infant brain cDNA library made by Dr. M. Soares

(Columbia University), oligo-dT primed and directionally

cloned between HindIII (5') and NotI (3') sites "

BASE COUNT 101 a 102 c 104 g 87 t 2 others

ORIGIN

Query Match 1.0%; Score 19; DB 14; Length 396;

Best Local Similarity 100.0%; Pred. No. 4.4e+02;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 370 AGCGTTCTTCGATCTT 388

Db 344 AGCGTTCTTCGATCTT 362

RESULT 36

LOCUS BJ529530 411 bp mRNA linear EST 09-AUG-2002

DEFINITION BJ529530 MF01SSB cDNA Oryzias latipes cDNA clone MF01SSB036B15 5',

mRNA sequence.

ACCSSION BJ529530

VERSION BJ529530.1 GI:22188342

KEYWORDS EST.

SOURCE Japanese medaka.

ORGANISM Oryzias latipes

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;

Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;

Belontiiformes; Adrianchthyidae; Oryziinae; Oryzias.

1 (bases 1 to 411)

Kohara,Y., Shin-i,T., Kimura,T., Narita,T., Jindo,T. and Takeda,H.

Medaka EST Project in Takeda's lab

Unpublished (2001)

Contact: Tadao Shin-i

Center For Genetic Resource Information

National Institute of Genetics

1111 Yata, Mishima, Shizuoka 411-8540, Japan

Tel: 81-559-81-6856

Fax: 81-559-81-6855

Email: tshin@genes.nig.ac.jp.

Location/Qualifiers

1. .411

/organism="Oryzias latipes"

/strain="Hd-R"

/db_xref="taxon:8090"

/clone_lib="MF01SSB036B15"

/clone_lib="MF01SSB cDNA"

/sex="mixture of female and male"

/tissue_type="whole embryo"

/dev_stage="segmentation stage 20 - 25"

80 a 104 c 99 g 128 t

Query Match 1.0%; Score 19; DB 13; Length 411;

Best Local Similarity 100.0%; Pred. No. 4.4e+02;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 849 TTTTAGGTTGCGTTCCTCT 867

Db 249 TTTTAGGTTGCGTTCCTCT 267

RESULT 37

LOCUS AA822992 419 bp mRNA linear EST 17-FEB-1998

DEFINITION AA822992 AA822992 5', mRNA sequence.

ACCSSION AA822992

VERSION AA822992.1 GI:2892860

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 419)

Marr,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,

Geisel,S., Kueba,T., Lacy,M., Le,M., Martin,J., Morris,M.,

Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
Theising, B., Wylie, F., Lennon, G., Soares, B., Wilson, R. and
Waterson, R.

TITLE JOURNAL COMMENT

The WashU-HHMI Mouse EST Project

Unpublished (1996)

Contact: Marra M/Mouse EST Project

WashU-HHMI Mouse EST Project

Washington University School of MedicineP

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@watson.wustl.edu

This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (infoimage.llnl.gov) for further information.

MG1:659877

Seq primer: -28m13 rev2 ET from Amersham.

FEATURES source

```

1. .419
Location/Qualifiers
/organism="Mus musculus"
/strain="C57BL/6J"
/db xref="taxon:10090"
/clone="IMAGE:1246189"
/clone_lib="Soares_mammary_gland_NbMMG"
/sex="male"
/tissue_type="mammary gland"
/dev_stages="4 weeks"
/lab_host="DH10B"
/note="Organ: mammary gland; Vector: pT7T3D-Pac (Pharmacia
RI; 1st strand cDNA was primed with a Not I - oligo(dT)
primer [5',
TGTACCAATCTGAAGTGGAGCGCCGCGAATGTTTTTTTTTTTTTTTTTTT
T 3']; Double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pT7T3 vector.
RNA provided by Dr. Minoru Ko, Wayne State Univ. Library
constructed and normalized by Bento Soares and M. Fatima
Bonaldo."
BASE COUNT      95 a      86 c      98 g      140 t
ORIGIN

```

Query Match 1.0%; Score 19; DB 9; Length 419;
Best Local Similarity 100.0%; Pred. NO. 4.5e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1645 CTTTAATTTTGCATTATT 1663

Db 236 CTTTAATTTTGCATTATT 254

RESULT 38 AZ290933

LOCUS AZ290933 430 bp DNA linear GSS 27-JUL-2000
DEFINITION RPCI-23-103A23.TJ RPCI-23 Mus musculus genomic clone RPCI-23-103A23
; DNA sequence.

ACCESSION AZ290933

VERSION AZ290933.1 GI:9532719

KEYWORDS GSS.

SOURCE house mouse.

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
1 (bases 1 to 430)
Zhao, S., Niemman, W., Feldblum, T., Malek, J., Shateman, S., Akinret
B., Levins, M., McGann, S., Teegaye, G., Geer, K., Krol, M., de Jong, P.
and Fraser, C.M.

TITLE Mouse BAC End Sequences from Library RPCI-23

JOURNAL Unpublished (1999)

COMMENT Contact: Shaying Zhao

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPCI-23. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (<http://bacpac.med.buffalo.edu/orderingframe.htm>)
or from Resea ch Genetics (info@resgen.com). BAC end page:
http://www.tigr.org/tdb/bac_end/mouse/bac_end_intro.html

Plate: 103 row: A column: 23

Seq primer: SP6

Class: BAC ends.

Location/Qualifiers

```

1. .430
/organism="Mus musculus"
/strain="C57BL/6J"
/db xref="taxon:10090"
/clone="RPCI-23-103A23"
/clone_lib="RPCI-23"
/sex="Female"
/lab_host="DH10B"
/note="Organ: Kidney/Brain; Vector: pBACe3.6; Site 1:
ECORI; Site 2: EORI; Female C57BL/6J mouse kidney and/or
brain genomic DNA was isolated and partially digested
with a combination of EORI and EORI Methyllase. Size
selected DNA was cloned into the pBACe3.6 vector at the
ECORI sites. The ligation products were transformed into
DH10B electrocompetent cells (BRL Life Technologies)."
BASE COUNT      139 a      66 c      91 g      134 t
ORIGIN

```

Query Match 1.0%; Score 19; DB 17; Length 430;

Best Local Similarity 100.0%; Pred. NO. 4.5e+02;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1712 CCACCTCTGTGTGTGTTTA 1730

Db 144 CCACCTCTGTGTGTGTTTA 162

RESULT 39

AQ359339/c

LOCUS

DEFINITION HS_5030_B2_D09_SP6E RPCI11 Human Male BAC Library Homo sapiens
genomic Clone Plate=606 Col=18 Row=H, DNA sequence.

ACCESSION AQ359339

VERSION AQ359339.1 GI:4208215

KEYWORDS GSS.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 430)

Mahairas, G.G., Wallace, J.C., Smith, K., Swartzell, S., Holzman, T., and
Keller, A., Shaker, R., Furlong, J., Young, J., Zhao, S., Adams, M.D., and
Hood, L.

Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome

Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)

JOURNAL

MEDLINE

COMMENT

Contact: Mahairas GG, Wallace JC, Hood L

High Throughput Sequencing Center

University of Washington

401 Queen Anne Avenue North, Seattle, WA 98109, USA

Tel: (206) 616-3618

Fax: (206) 616-3887

Email: jwallace@u.washington.edu

Clones may be purchased from Research Genetics (info@resgen.com).

BAC end Web Server: <http://www.htsc.washington.edu>

Plate: 606 row: H column: 18

Seq primer: SP6

Class: BAC ends

High quality sequence stop: 430.

Location/Qualifiers

```

source
1. .430
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=606 Col=18 Row=H"
/clone_1lb="RPC111 Human Male BAC Library"
/sex="Male"
/cell_type="lymphocytes"
/note="vector: pBAC3.6; RPC111 Human Male BAC Library"

BASE COUNT      113 a      91 c      93 g      132 t      1 others
ORIGIN
Query Match      1.0%; Score 19; DB 17; Length 430;
Best Local Similarity 100.0%; Pred. No. 4.5e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1151 ATTGTTGTCAGAACAGGC 1169
|||||
Db 245 ATTGTTGTCAGAACAGGC 227

RESULT 40
BH864257/c 447 bp DNA linear GSS 05-AUG-2002
LOCUS      SALK_095658 Arabidopsis thaliana TDNA insertion lines Arabidopsis
DEFINITION thaliana genomic clone SALK_095658, DNA sequence.
ACCESSION  BH864257
VERSION     BH864257.1 GI:22100155
KEYWORDS   GSS.
SOURCE     thale cress.
ORGANISM   Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
            1 (bases 1 to 447)
            Alonso,J.M., Leises,T.J., Barajas,P., Chen,H., Cheuk,R., Gadrinab
            ,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L., Shinn,P.
            , Zimmerman,J. and Ecker,J.R.
            A Sequence-Indexed Library of Insertion Mutations in the
            Arabidopsis Genome
            Unpublished (2001)
            Contact: Joseph R. Ecker
            Salk Institute Genomic Analysis Laboratory (SIGAL)
            10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
            Tel: 858 453 4100 x1752
            Fax: 858 558 6379
            Email: ecker@salk.edu
            This is single pass sequence recovered from the left border of
            TDNA. This sequence lies within an annotated exon of Atlg23010.
            Class: TDNA tagged.
            Location/Qualifiers
            1. .447
            /organism="Arabidopsis thaliana"
            /strain="Columbia 0"
            /db_xref="taxon:3702"
            /clone="SALK_095658"
            /note="PCR was performed on Arabidopsis thaliana lines
            each of which contains one or more TDNA insertion
            elements. The resultant fragment for each line was
            directly sequenced to determine the genomic sequence at
            the site of insertion. Details of the protocols used can
            be found at http://signal.salk.edu/cdna_protocols.html"

BASE COUNT      104 a      104 c      114 g      125 t
ORIGIN
Query Match      1.0%; Score 19; DB 17; Length 447;
Best Local Similarity 100.0%; Pred. No. 4.5e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1478 TTGATCATTTTGGATCGTA 1496
|||||
Db 163 TTGATCATTTTGGATCGTA 145

RESULT 41
AV968816/c 455 bp mRNA linear EST 14-MAR-2002
LOCUS      AV968816 Nori Satoh unpublished cDNA library, larva Ciona
DEFINITION intestinalis cDNA clone c1v16p19 5', mRNA sequence.
ACCESSION  AV968816
VERSION     AV968816.1 GI:19458580
KEYWORDS   EST.
SOURCE     Ciona intestinalis.
ORGANISM   Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
            Phlebobranchia; Clonidae; Ciona.
            1 (bases 1 to 455)
            Satoh,N., Satou,Y., Kohara,Y. and Shin-I,T.
            Expressed genes in Ciona intestinalis
            Unpublished (2000)
            Contact: Nori Satoh
            Department of Zoology
            Kyoto University
            Sakyo-Ku, Kyoto, Kyoto 606-8502, Japan
            Tel: 81-75-753-4081
            Fax: 81-75-705-1113
            Email: satoh@ascidian.zool.kyoto-u.ac.jp.
            Location/Qualifiers
            1. .455
            /organism="Ciona intestinalis"
            /db_xref="taxon:7719"
            /clone="c1v16p19"
            /clone_1lb="Nori Satoh unpublished cDNA library, larva"
            /tissue_type="whole animal"
            /dev_stage="larva"
            /note="vector: pBluescript SK"

BASE COUNT      149 a      77 c      113 g      113 t      3 others
ORIGIN
Query Match      1.0%; Score 19; DB 10; Length 455;
Best Local Similarity 100.0%; Pred. No. 4.5e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 210 CGATGCTTTTGTCCCGA 228
|||||
Db 354 CGATGCTTTTGTCCCGA 336

RESULT 42
AA798200 459 bp mRNA linear EST 10-FEB-1998
LOCUS      ub60e03.r1 Soares mammary_gland NMLMG Mus musculus cDNA clone
DEFINITION IMAGE:1382140 5' similar to TR:O10384 O10384 GAG POLYPROTEIN. [1]
            ; mRNA sequence.
            1 (bases 1 to 459)
            Soares,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theisinger,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
            The WashU-HMI Mouse EST Project
            Unpublished (1996)
            Contact: Marra M/Mouse EST Project
            WashU-HMI Mouse EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810

```

Email: mouseest@watson.wustl.edu
 This clone is available royalty-free through LLNL ; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 MGI:904608
 Seq primer: -28ml3 rev2 ET from Amersham
 High quality sequence stop: 451.
 Location/Qualifiers

FEATURES

source

1..459
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="IMAGE:1382140"
 /clone_lib="Soares mammary_gland_NMLMG"
 /sex="female (lactating)"
 /tissue_type="mammary gland"
 /lab_host="DH10B"

/note="vector: pT7T3D-Pac (Pharmacia) with a modified
 polylinker; 1st strand cDNA was prepared from mammary
 gland tissue from a lactating female, and was then primed
 with a Not I - oligo(dT) primer. Double-stranded cDNA was
 ligated to Eco RI adaptors (Pharmacia), digested with Not
 I and cloned into the Not I and Eco RI sites of the
 modified pT7T3 vector. Library is normalized. Library
 was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT, 137 a 106 c 92 g 124 t

ORIGIN

Query Match 1.0%; Score 19; DB 9; Length 459;
 Best Local Similarity 100.0%; Pred. No. 4.5e+02;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 894 TCGGAATCGATTTTAAAGG 912

Db 421 TCGGAATCGATTTTAAAGG 439

RESULT 43

LOCUS AQ698248 459 bp DNA linear GSS 06-JUL-1999
 DEFINITION HS 5550_B2_E04_SP6E RPCI-11 Human Male BAC Library Homo sapiens
 genomic clone Plates=1126 Col=8 Row=J, DNA sequence.

ACCESSION AQ698248

VERSION AQ698248.1 GI:5388496

KEYWORDS GSS.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
 Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
 Hood,L.

TITLE Sequence-tagged connectors: A sequence approach to mapping and

scanning the human genome

Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)

JOURNAL 99380589

MEDLINE

COMMENT

Contact: Mahairas GG, Wallace JC, Hood L

High Throughput Sequencing Center

University of Washington

401 Queen Anne Avenue North, Seattle, WA 98109, USA

Tel: (206) 616-3618

Fax: (206) 616-3887

Email: jwallace@u.washington.edu

Clones are derived from the human BAC library RPCI-11. For BAC
 library availability, please contact Pieter de Jong

(pieter@dejong.med.buffalo.edu). Clones may be purchased from

BACPAC Resources (http://bacpac.med.buffalo.edu/ordering.bac.htm)

or from Research Genetics (info@resgen.com). BAC end web server:

http://www.htsc.washington.edu

Plate: 1126 row: J column: 8

Seq primer: SP6

Class: BAC ends

High quality sequence stop: 459.

Location/Qualifiers

FEATURES

source

1..459
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="Plate=1126 Col=8 Row=J"
 /clone_lib="RPCI-11 Human Male BAC Library"
 /sex="male"
 /note="vector: pBACE3.6; Site 1: EcoRI; Site 2: EcoRI;
 Male blood DNA was isolated from one randomly chosen donor
 and partially digested with a combination of EcoRI and
 EcoRI Methylase. Size selected DNA was cloned into the
 pBACE3.6 vector at EcoRI sites"

BASE COUNT 118 a 81 c 104 g 156 t

ORIGIN

Query Match 1.0%; Score 19; DB 17; Length 459;
 Best Local Similarity 100.0%; Pred. No. 4.5e+02;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 511 AGAGGAATTTTATGTCT 529

Db 310 AGAGGAATTTTATGTCT 328

RESULT 44

LOCUS A1159020

DEFINITION A1159020 461 bp mRNA linear EST 02-OCT-1998
 v286f10.r1 Soares_mammary_gland_NbMMG Mus musculus cDNA clone
 IMAGE:1333387 5', mRNA sequence.

ACCESSION A1159020

VERSION A1159020.1 GI:3692202

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

AUTHORS

Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
 Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
 Scheilenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
 Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
 Waterston,R.

TITLE The WashU-HMMI Mouse EST Project

Unpublished (1996)

Contact: Marra M/Mouse EST Project

WashU-HMMI Mouse EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@watson.wustl.edu

This clone is available royalty-free through LLNL ; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

MGI:692931

Seq primer: -28ml3 rev2 ET from Amersham

High quality sequence stop: 436.

Location/Qualifiers

source

1..461
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="IMAGE:1333387"
 /clone_lib="Soares_mammary_gland_NbMMG"
 /sex="male"
 /tissue_type="mammary gland"
 /dev_stage="4 weeks"
 /lab_host="DH10B"

/note="Organ: mammary gland; Vector: pT7T3D-Pac (Pharmacia
) with a modified polylinker; Site 1: Not I; Site 2: Eco
 RI; 1st strand cDNA was primed with a Not I - oligo(dT)

primer [5'

TGTTACCATCTCAAGTGGAGCGCGCAATGTTTTTTTTTTTTTTTTTTTTT
 T 3']; double-stranded cDNA was ligated to Eco RI

adaptors (Pharmacia), digested with Not I and cloned into

the Not I and Eco RI sites of the modified pRT73 vector.
RNA provided by Dr. Minoru Ko, Wayne State Univ. Library
constructed and normalized by Bento Soares and M. Patricia
Bonaldo."

BASE COUNT 127 a 78 c 136 g 120 t
ORIGIN
Query Match 1.0%; Score 19; DB 9; Length 461;
Best Local Similarity 100.0%; Pred. No. 4.5e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 1645 CTTTAATTTTCATTATT 1663
Db 381 CTTTAATTTTCATTATT 399

RESULT 45 481 bp DNA linear GSS 04-OCT-2000
AZ463779
LOCUS 1M0272A19R Mouse 10kb plasmid UGCLM library Mus musculus genomic
DEFINITION
clone UGCLM0272A19 R, DNA sequence.

ACCESSION AZ463779
VERSION AZ463779.1 GI:10621904
KEYWORDS GSS.

SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 481)
Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,
Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly,
'M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausen, A.
and Wright, D., Weiss, R.

AUTHORS Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
Unpublished (2000)
CONTACT: Robert B. Weiss
JOURNAL University of Utah Genome Center
COMMENT Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA

Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0272 row: A column: 19
Seq primer: CACACAGAAACGCTATGACC
Class: plasmid ends
High quality sequence stop: 481.

FEATURES
source Location/Qualifiers
1. 481

/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UGCLM0272A19"
/clone_lib="Mouse 10kb plasmid UGCLM library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, TI-resistant, F-"
/note="Vector: pMD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adapted DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of pMD42 (GII4732114[GB|AF129072.1]), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and

purified. The sheared, adapted mouse DNA was annealed to
adapted vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."

BASE COUNT 193 a 85 c 74 g 129 t
ORIGIN
Query Match 1.0%; Score 19; DB 17; Length 481;
Best Local Similarity 100.0%; Pred. No. 4.5e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 1460 TATTCATTAACATACCTT 1478
Db 458 TATTCATTAACATACCTT 476

Search completed: December 17, 2002, 10:34:23
Job time : 2058 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 17, 2002, 10:34:31 , Search time 73 Seconds
(without alignments)
10097.799 Million cell updates/sec

Title: US-09-868-987-1
Perfect score: 1864
Sequence: 1 atgacttcgcataatgtc.....atgacaaattcagataatgc 1864

Scoring table: OIIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 355320 seqs, 197730502 residues

Word size : 0
Total number of hits satisfying chosen parameters: 710640

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : Published Applications NA.*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq.*
11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq.*
13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query			ID	Description
	Score	Match	Length		
1	1864	100.0	1864	10	US-09-886-468-1
2	20	1.1	448	10	US-09-983-965-4080
3	19	1.0	439	10	US-09-864-761-10333
C 4	19	1.0	445	10	US-09-864-761-26966
C 5	19	1.0	469	10	US-09-864-761-32700
C 6	19	1.0	561	10	US-09-864-761-16209
7	19	1.0	1150	10	US-09-822-849A-374
8	19	1.0	1707	9	US-09-938-842A-186
9	19	1.0	2000	9	US-09-938-842A-4668
C 10	19	1.0	3470	12	US-10-044-090-516
11	18	1.0	366	10	US-09-878-574-2011
12	18	1.0	392	10	US-09-924-035A-470
C 13	18	1.0	421	10	US-09-983-965-2819
C 14	18	1.0	459	10	US-09-920-300A-17
C 15	18	1.0	459	12	US-10-033-528-17
C 16	18	1.0	478	10	US-09-770-444-3
17	18	1.0	613	10	US-09-770-149-852
18	1.0	713	10	US-09-764-847-1360	
19	18	1.0	755	9	US-09-938-842A-5222
					Sequence 1, Appli
					Sequence 4080, Ap
					Sequence 10333, Ap
					Sequence 26966, A
					Sequence 32700, A
					Sequence 16209, A
					Sequence 374, App
					Sequence 186, App
					Sequence 4668, Ap
					Sequence 516, App
					Sequence 2011, Ap
					Sequence 470, App
					Sequence 2819, Ap
					Sequence 17, Appl
					Sequence 3, Appli
					Sequence 852, App
					Sequence 1360, Ap
					Sequence 5222, Ap

Sequence 715, App
Sequence 1361, Ap
Sequence 2071, Ap
Sequence 2072, Ap
Sequence 3, Appli
Sequence 26319, A
Sequence 1387, Ap
Sequence 1387, Ap
Sequence 452, App
Sequence 6153, Ap
Sequence 3885, Ap
Sequence 5590, Ap
Sequence 1850, Ap
Sequence 10630, A
Sequence 67, Appl
Sequence 7444, Ap
Sequence 65, Appl
Sequence 157, App
Sequence 9, Appli
Sequence 161, App
Sequence 5070, Ap
Sequence 9546, Ap
Sequence 116, App
Sequence 173, App
Sequence 1, Appli
Sequence 3, Appli
Sequence 4722, Ap
Sequence 8636, Ap
Sequence 6, Appli
Sequence 1582, Ap
Sequence 1586, Ap
Sequence 1584, Ap
Sequence 1583, Ap
Sequence 1, Appli
Sequence 46, Appl
Sequence 46, Appl
Sequence 2184, Ap
Sequence 3, Appli
Sequence 3, Appli
Sequence 3, Appli
Sequence 1, Appli
Sequence 99, Appl
Sequence 29127, A
Sequence 172, App
Sequence 32118, A
Sequence 3542, Ap
Sequence 967, App
Sequence 1319, Ap
Sequence 13252, A
Sequence 6487, Ap
Sequence 536, App
Sequence 1539, Ap
Sequence 167, App
Sequence 112, App
Sequence 1194, Ap
Sequence 222, App
Sequence 436, App
Sequence 267, App
Sequence 1863, Ap
Sequence 11588, A
Sequence 65, Appl
Sequence 489, App
Sequence 439, App
Sequence 8773, Ap
Sequence 4878, Ap
Sequence 14446, A
Sequence 1078, Ap
Sequence 9849, Ap
Sequence 4026, Ap
Sequence 7924, Ap
Sequence 264, App
Sequence 15066, A
Sequence 1051, Ap

531	15	0.8	750	10	US-09-974-300-6310	Sequence 6310, Ap	Sequence 6310, Ap	15	0.8	1323	10	US-09-778-320-38	Sequence 38, Appl
c 532	15	0.8	753	10	US-09-970-300-7239	Sequence 7239, Ap	Sequence 7239, Ap	15	0.8	1323	10	US-09-910-689-38	Sequence 38, Appl
c 533	15	0.8	768	10	US-09-910-943-187	Sequence 187, Ap	Sequence 187, Ap	15	0.8	1323	12	US-10-010-742-38	Sequence 38, Appl
c 534	15	0.8	771	9	US-09-974-300-2926	Sequence 2926, Ap	Sequence 2926, Ap	15	0.8	1325	10	US-09-765-205-1	Sequence 1, Appl
c 535	15	0.8	773	9	US-09-771-035A-52	Sequence 52, Appl	Sequence 52, Appl	15	0.8	1333	9	US-09-978-295A-321	Sequence 321, Appl
c 536	15	0.8	789	9	US-09-813-453A-31	Sequence 31, Appl	Sequence 31, Appl	15	0.8	1333	9	US-09-978-697-321	Sequence 321, Appl
c 537	15	0.8	813	10	US-09-770-445-763	Sequence 763, Ap	Sequence 763, Ap	15	0.8	1333	9	US-09-978-697-321	Sequence 321, Appl
c 538	15	0.8	836	10	US-09-770-445-675	Sequence 675, Ap	Sequence 675, Ap	15	0.8	1333	9	US-09-978-697-321	Sequence 321, Appl
c 539	15	0.8	852	10	US-09-770-445-623	Sequence 623, Ap	Sequence 623, Ap	15	0.8	1344	9	US-09-938-842A-2666	Sequence 2666, Ap
c 540	15	0.8	854	10	US-09-770-445-618	Sequence 618, Ap	Sequence 618, Ap	15	0.8	1344	9	US-09-938-842A-4296	Sequence 4296, Ap
c 541	15	0.8	877	9	US-09-938-842A-5040	Sequence 5040, Ap	Sequence 5040, Ap	15	0.8	1345	9	US-09-938-842A-5125	Sequence 5125, Ap
c 542	15	0.8	891	10	US-09-815-242-7115	Sequence 7115, Ap	Sequence 7115, Ap	15	0.8	1350	9	US-09-895-913A-291	Sequence 291, Appl
c 543	15	0.8	893	10	US-09-764-878-277	Sequence 277, Ap	Sequence 277, Ap	15	0.8	1356	10	US-09-815-242-6624	Sequence 6624, Ap
c 544	15	0.8	893	10	US-09-764-878-277	Sequence 277, Ap	Sequence 277, Ap	15	0.8	1362	9	US-09-974-298-73	Sequence 73, Appl
c 545	15	0.8	894	10	US-09-764-860-1143	Sequence 1140, Ap	Sequence 1140, Ap	15	0.8	1362	12	US-10-044-090-556	Sequence 556, Appl
c 546	15	0.8	894	10	US-09-760-927A-683	Sequence 683, Ap	Sequence 683, Ap	15	0.8	1371	10	US-09-801-368-203	Sequence 203, Appl
c 547	15	0.8	912	9	US-09-764-868-272	Sequence 272, Ap	Sequence 272, Ap	15	0.8	1377	10	US-09-801-368-167	Sequence 167, Appl
c 548	15	0.8	912	9	US-09-764-868-581	Sequence 581, Ap	Sequence 581, Ap	15	0.8	1386	10	US-09-974-300-1151	Sequence 1151, Appl
c 549	15	0.8	930	10	US-09-815-242-7193	Sequence 7193, Ap	Sequence 7193, Ap	15	0.8	1389	9	US-09-938-842A-340	Sequence 340, Appl
c 550	15	0.8	936	10	US-09-815-242-4764	Sequence 4764, Ap	Sequence 4764, Ap	15	0.8	1389	9	US-09-938-842A-340	Sequence 340, Appl
c 551	15	0.8	960	9	US-09-938-842A-1228	Sequence 12, Appl	Sequence 12, Appl	15	0.8	1391	12	US-10-044-477-2	Sequence 2, Appl
c 552	15	0.8	972	10	US-09-815-242-6142	Sequence 6142, Ap	Sequence 6142, Ap	15	0.8	1404	10	US-09-729-835-34	Sequence 34, Appl
c 553	15	0.8	978	10	US-09-815-242-3800	Sequence 3800, Ap	Sequence 3800, Ap	15	0.8	1428	10	US-09-741-669-265	Sequence 265, Appl
c 554	15	0.8	978	10	US-09-815-242-8849	Sequence 8849, Ap	Sequence 8849, Ap	15	0.8	1431	10	US-09-774-414-2	Sequence 2, Appl
c 555	15	0.8	999	10	US-09-815-242-8944	Sequence 8944, Ap	Sequence 8944, Ap	15	0.8	1440	9	US-09-938-842A-80	Sequence 80, Appl
c 556	15	0.8	1003	9	US-09-938-842A-4517	Sequence 4517, Ap	Sequence 4517, Ap	15	0.8	1462	10	US-09-220-091-6	Sequence 6, Appl
c 557	15	0.8	1034	9	US-09-736-457-1654	Sequence 1654, Ap	Sequence 1654, Ap	15	0.8	1467	9	US-09-938-842A	


```
C 969 14 0.8 62 10 US-09-878-574-9009 Sequence 9009, Ap
970 14 0.8 75 10 US-09-864-761-24218 Sequence 24218, A
971 14 0.8 76 10 US-09-864-761-23727 Sequence 23727, A
C 972 14 0.8 80 10 US-09-864-761-28939 Sequence 28939, A
973 14 0.8 91 9 US-10-035-098-24 Sequence 24, Appl
974 14 0.8 92 12 US-10-029-913-3 Sequence 3, Appl
C 975 14 0.8 97 9 US-10-035-098-23 Sequence 23, Appl
976 14 0.8 100 10 US-09-864-761-31448 Sequence 31448, A
C 977 14 0.8 101 10 US-09-867-701-2914 Sequence 2914, Ap
978 14 0.8 102 10 US-09-864-761-32561 Sequence 32561, A
C 979 14 0.8 103 9 US-10-046-935-1608 Sequence 1608, Ap
980 14 0.8 103 9 US-09-878-178-1608 Sequence 1608, Ap
981 14 0.8 106 10 US-09-969-373-1342 Sequence 1342, Ap
982 14 0.8 108 10 US-09-864-761-22137 Sequence 22137, A
983 14 0.8 111 10 US-09-864-761-23456 Sequence 23456, A
984 14 0.8 114 10 US-09-815-242-7020 Sequence 7020, Ap
985 14 0.8 114 10 US-09-867-701-1549 Sequence 1549, Ap
C 986 14 0.8 118 10 US-09-864-761-16867 Sequence 16867, A
987 14 0.8 124 10 US-09-864-761-30335 Sequence 30335, A
988 14 0.8 127 10 US-09-864-761-29853 Sequence 29853, A
C 989 14 0.8 129 10 US-09-923-876-5782 Sequence 5782, Ap
C 990 14 0.8 129 10 US-09-864-761-19910 Sequence 19910, A
991 14 0.8 131 10 US-09-867-701-2673 Sequence 2673, Ap
C 992 14 0.8 132 10 US-09-864-761-26847 Sequence 26847, A
C 993 14 0.8 136 10 US-09-867-701-5102 Sequence 5102, Ap
C 994 14 0.8 138 10 US-09-974-300-7573 Sequence 7573, Ap
995 14 0.8 141 10 US-09-783-590-9503 Sequence 9503, Ap
C 996 14 0.8 144 10 US-09-770-696-525 Sequence 525, App
C 997 14 0.8 148 10 US-09-864-761-21728 Sequence 21728, A
998 14 0.8 148 10 US-09-864-761-17911 Sequence 17911, A
999 14 0.8 149 10 US-09-864-761-30530 Sequence 30530, A
1000 14 0.8 149 10 US-09-969-373-316 Sequence 316, App
```

ALIGNMENTS

```
RESULT 1
US-09-886-468-1
; Sequence 1, Application US/09886468
; Patent No. US20020037293A1
; GENERAL INFORMATION:
; APPLICANT: Aventis Pasteur Limited
; TITLE OF INVENTION: Chlamydia antigens and corresponding DNA fragments and uses there
; FILE REFERENCE: 77813-5
; CURRENT APPLICATION NUMBER: US/09/886,468
; PRIOR FILING DATE: 1999-12-23
; PRIOR APPLICATION NUMBER: 60/113,280
; PRIOR FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: 60/113,281
; PRIOR FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: 60/113,282
; PRIOR FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: 60/113,283
; PRIOR FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: 60/113,284
; PRIOR FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: 60/113,285
; PRIOR FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: 60/113,385
; PRIOR FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: 60/114,050
; PRIOR FILING DATE: 1998-12-28
; PRIOR APPLICATION NUMBER: 60/114,056
; PRIOR FILING DATE: 1998-12-28
; PRIOR APPLICATION NUMBER: 60/114,057
; PRIOR FILING DATE: 1998-12-28
; PRIOR APPLICATION NUMBER: 60/114,058
; PRIOR FILING DATE: 1998-12-28
; PRIOR APPLICATION NUMBER: 60/114,059
; PRIOR FILING DATE: 1998-12-28
; PRIOR APPLICATION NUMBER: 60/114,061
; PRIOR FILING DATE: 1998-12-28
```

```
NUMBER OF SEQ ID NOS: 26
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1
; LENGTH: 1864
; TYPE: DNA
; ORGANISM: Chlamydia pneumoniae
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (101)..(1756)
US-09-886-468-1
Query Match 100.0%; Score 1864; DB 10; Length 1864;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1864; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGAGCTTCGCAATATGTCAGAGGGGATCAGCGGCACTGCTAATGACAAATATTCG 60
1 ATGAGCTTCGCAATATGTCAGAGGGGATCAGCGGCACTGCTAATGACAAATATTCG 60
DB 1
QY 61 AAACCTGGATGGCGATGCTGTATGATGACGGTTATATGTCAGACGCCCTATTTT 120
61 AAACCTGGATGGCGATGCTGTATGATGACGGTTATATGTCAGACGCCCTATTTT 120
DB 61 AAACCTGGATGGCGATGCTGTATGATGACGGTTATATGTCAGACGCCCTATTTT 120
121 AAACCTGGATGGCGATGCTGTATGATGACGGTTATATGTCAGACGCCCTATTTT 180
121 AAACCTGGATGGCGATGCTGTATGATGACGGTTATATGTCAGACGCCCTATTTT 180
DB 121 AAACCTGGATGGCGATGCTGTATGATGACGGTTATATGTCAGACGCCCTATTTT 180
181 CAAACTGCGCTCAGATTTAAATCTGGACGATGCTTTTGTCCGAGGTTCTCAGTA 240
181 CAAACTGCGCTCAGATTTAAATCTGGACGATGCTTTTGTCCGAGGTTCTCAGTA 240
DB 181 CAAACTGCGCTCAGATTTAAATCTGGACGATGCTTTTGTCCGAGGTTCTCAGTA 240
241 AGAGAGATCTCTTCGATCTTGGGAAAAACAATGACAAAGGATATCTCAGCAG 300
241 AGAGAGATCTCTTCGATCTTGGGAAAAACAATGACAAAGGATATCTCAGCAG 300
DB 241 AGAGAGATCTCTTCGATCTTGGGAAAAACAATGACAAAGGATATCTCAGCAG 300
301 CTGTGCTTGGCAATGCTATGTTTGGATGAGCGTATATATGATTTGAGGCGCTCAT 360
301 CTGTGCTTGGCAATGCTATGTTTGGATGAGCGTATATATGATTTGAGGCGCTCAT 360
DB 301 CTGTGCTTGGCAATGCTATGTTTGGATGAGCGTATATATGATTTGAGGCGCTCAT 360
361 CGCTTGGGAGCGTCTTCTGAACTTTTGTCTTATCTGAGGCGCTTACAGTATTTGA 420
361 CGCTTGGGAGCGTCTTCTGAACTTTTGTCTTATCTGAGGCGCTTACAGTATTTGA 420
DB 361 CGCTTGGGAGCGTCTTCTGAACTTTTGTCTTATCTGAGGCGCTTACAGTATTTGA 420
421 TGGCCACTCAGCTTCTGAGACTCGCTGGATGTTCTTGCATAGGAGTGGCCCTA 480
421 TGGCCACTCAGCTTCTGAGACTCGCTGGATGTTCTTGCATAGGAGTGGCCCTA 480
DB 421 TGGCCACTCAGCTTCTGAGACTCGCTGGATGTTCTTGCATAGGAGTGGCCCTA 480
481 TGCAAATGTTCTGTATGAAAGATCCAGAGGATTTTATGTCCTCAAGCTTAA 540
481 TGCAAATGTTCTGTATGAAAGATCCAGAGGATTTTATGTCCTCAAGCTTAA 540
DB 481 TGCAAATGTTCTGTATGAAAGATCCAGAGGATTTTATGTCCTCAAGCTTAA 540
541 AAATCTGTAGAAAAGATATCAAGGCTTTGGAGCCATTTTGAATCTTAACTTGAC 600
541 AAATCTGTAGAAAAGATATCAAGGCTTTGGAGCCATTTTGAATCTTAACTTGAC 600
DB 541 AAATCTGTAGAAAAGATATCAAGGCTTTGGAGCCATTTTGAATCTTAACTTGAC 600
601 TACAGATTTGGCTCAGCACTTCTTCTCCAGATACAGGCGCTTAAAGGTTTGC 660
601 TACAGATTTGGCTCAGCACTTCTTCTCCAGATACAGGCGCTTAAAGGTTTGC 660
DB 601 TACAGATTTGGCTCAGCACTTCTTCTCCAGATACAGGCGCTTAAAGGTTTGC 660
661 TTGACATTTGATTTGGAATTTCTCTCAATGTTTACGCGCTTTGATGACTAAAT 720
661 TTGACATTTGATTTGGAATTTCTCTCAATGTTTACGCGCTTTGATGACTAAAT 720
DB 661 TTGACATTTGATTTGGAATTTCTCTCAATGTTTACGCGCTTTGATGACTAAAT 720
721 TTTCCTCATGCTGTGATGATTAAGCCCAACATACAGTTGATGATGATTAAGTT 780
721 TTTCCTCATGCTGTGATGATTAAGCCCAACATACAGTTGATGATGATTAAGTT 780
DB 721 TTTCCTCATGCTGTGATGATTAAGCCCAACATACAGTTGATGATGATTAAGTT 780
781 CGTGGGATTAAGCATGATTTCTTGAGAGATGCAAAAATTTGGGCTGTTCTGAAG 840
781 CGTGGGATTAAGCATGATTTCTTGAGAGATGCAAAAATTTGGGCTGTTCTGAAG 840
DB 781 CGTGGGATTAAGCATGATTTCTTGAGAGATGCAAAAATTTGGGCTGTTCTGAAG 840
841 TGTTCCTTTAGTGTGCTGCTCGGGGTTTGGAGCTGGAATTCGTTTGGGAAT 900
841 TGTTCCTTTAGTGTGCTGCTCGGGGTTTGGAGCTGGAATTCGTTTGGGAAT 900
```

Db 841 TGTCTTTCTTTAGTTCGTTGCTCTCGGTTTGGAGCCTCGAATTCGGTTTGGGAAT 900
Qy 901 GGATTTTAAAGAGGGTATGCTTTTACCTTTAATCAAAAGAGCATGGCATCAGCGATGT 960
Db 901 GGATTTTAAAGAGGGTATGCTTTTACCTTTAATCAAAAGAGCATGGCATCAGCGATGT 960
Qy 961 TGCTCAAAATGGTGGCAAAAGTTGTGCATAAATACAGGAAGCTGCTTTCTTTAGAGA 1020
Db 961 TGCTCAAAATGGTGGCAAAAGTTGTGCATAAATACAGGAAGCTGCTTTCTTTAGAGA 1020
Qy 1021 CTTCGGTATTCAAAATTTGGATCTTCAGAAAAGATCAAAATCTATTATTAGTATAAAGC 1080
Db 1021 CTTCGGTATTCAAAATTTGGATCTTCAGAAAAGATCAAAATCTATTATTAGTATAAAGC 1080
Qy 1081 TTTAAGCTATAAGAGAGATACAGAGCTCTCTCTAAATTAACGATCATGAGCTGGCG 1140
Db 1081 TTTAAGCTATAAGAGAGATACAGAGCTCTCTCTAAATTAACGATCATGAGCTGGCG 1140
Qy 1141 TTATTGTGGGATTGTTGTGCAGAAAAGAGCTAGATTCTCTACGGAAATCTTAAACGAAA 1200
Db 1141 TTATTGTGGGATTGTTGTGCAGAAAAGAGCTAGATTCTCTACGGAAATCTTAAACGAAA 1200
Qy 1201 CGCAAAATTTTGGTCAAAAGTTAAGCAGCAAACTATCGAAGAAAATGCGTTATCAGCGCAC 1260
Db 1201 CGCAAAATTTTGGTCAAAAGTTAAGCAGCAAACTATCGAAGAAAATGCGTTATCAGCGCAC 1260
Qy 1261 CATCGGCTTTTGGAGCTTTGGCAATCATCTGCTATGATGAGTTTGGCTGCGTAAATG 1320
Db 1261 CATCGGCTTTTGGAGCTTTGGCAATCATCTGCTATGATGAGTTTGGCTGCGTAAATG 1320
Qy 1321 GCAATATGCTTTCAAGTGGCGTATGCGCTTTTAAATCATGACCTTTTGGCTACCTGTGCGAT 1380
Db 1321 GCAATATGCTTTCAAGTGGCGTATGCGCTTTTAAATCATGACCTTTTGGCTACCTGTGCGAT 1380
Qy 1381 CTGTGTTTATAGCACATTTCTTTTCAAGAAAATTCAAATAGATTTTGCAGCCATTGGTGC 1440
Db 1381 CTGTGTTTATAGCACATTTCTTTTCAAGAAAATTCAAATAGATTTTGCAGCCATTGGTGC 1440
Qy 1441 TTTAATGATGTTATGGGATATTCATTAACCAATATCTTTGATCAITTTTGATCGTATTTCG 1500
Db 1441 TTTAATGATGTTATGGGATATTCATTAACCAATATCTTTGATCAITTTTGATCGTATTTCG 1500
Qy 1501 TGAAGATCGCAAGCGAACCTGTTTACCCCTATGATGCTTTTAGTTAATGATGCCCTTCA 1560
Db 1501 TGAAGATCGCAAGCGAACCTGTTTACCCCTATGATGCTTTTAGTTAATGATGCCCTTCA 1560
Qy 1561 AAAGACGTTTCAGCCGACGTTAATGACAAACAGCTACAATCTATCAGTTTGTGTTAATGCT 1620
Db 1561 AAAGACGTTTCAGCCGACGTTAATGACAAACAGCTACAATCTATCAGTTTGTGTTAATGCT 1620
Qy 1621 TTTGTTTATAGCGGCTCCTCTGCTTTTAAATTTTGCATTTATATGACCATAGGATTC 1680
Db 1621 TTTGTTTATAGCGGCTCCTCTGCTTTTAAATTTTGCATTTATATGACCATAGGATTC 1680
Qy 1681 TCTAGAACTTTATGCTCTCTTTATTTGACCACTCTGTTGTTTATGTTATGTCGGTAA 1740
Db 1681 TCTAGAACTTTATGCTCTCTTTATTTGACCACTCTGTTGTTTATGTTATGTCGGTAA 1740
Qy 1741 AGAAAATCGCTCAAAATAGTACCGTTAACTTAATCTAACGTTAGCAATATAAAATC 1800
Db 1741 AGAAAATCGCTCAAAATAGTACCGTTAACTTAATCTAACGTTAGCAATATAAAATC 1800
Qy 1801 TCCTTTGGGACTTTAGTCCCAAGGCCCTGTGGTATTAAATTTATGACAAATTCAGATA 1860
Db 1801 TCCTTTGGGACTTTAGTCCCAAGGCCCTGTGGTATTAAATTTATGACAAATTCAGATA 1860
Qy 1861 ATGC 1864
Db 1861 ATGC 1864

RESULT 2
US-09-983-965-4080

; Sequence 4080, Application US/09983965
; Patent No. US20020137160A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Mengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION
; FILE REFERENCE: 37-21(10297)C
; CURRENT APPLICATION NUMBER: US/09/983,965
; PRIOR FILING DATE: 2001-10-26
; PRIOR APPLICATION NUMBER: US 09/465,231
; PRIOR FILING DATE: 1999-12-15
; PRIOR APPLICATION NUMBER: US 60/113,678
; PRIOR FILING DATE: 1998-12-17
; NUMBER OF SEQ ID NOS: 5912
; SEQ ID NO 4080
; LENGTH: 448
; TYPE: DNA
; ORGANISM: Bos taurus
; FEATURE:
; OTHER INFORMATION: Clone ID: 56-LIB3058-005-Q1-K1-F8
US-09-983-965-4080

Query Match 1.1%; Score 20; DB 10; Length 448;
Best Local Similarity 100.0%; Pred. No. 1.8;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1641 CTGCTCTTAATTTTGCATTT 1660
Db 357 CTGCTCTTAATTTTGCATTT 376

RESULT 3
US-09-864-761-10333/C
; Sequence 10333, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662

```

; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 10333
; LENGTH: 439
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC012460.2
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.98
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1
US-09-864-761-10333

Query Match          1.0%; Score 19; DB 10; Length 439;
Best Local Similarity 100.0%; Pred. No. 6;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 370 AGCTGTTCTTCTGATCTT 388
DB 412 AGCTGTTCTTCTGATCTT 394

RESULT 4
US-09-864-761-26966/c
; Sequence 26966, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecm1ca-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
```

```

; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 26966
; LENGTH: 445
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC012460.2
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.98
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1
; OTHER INFORMATION: EST HUMAN HIT: BE264998.1, EVALU0 0.00e+00
; OTHER INFORMATION: SWISSPROT HIT: P34384, EVALU0 1.50e-02
; OTHER INFORMATION: NT HIT: AE002179.2, EVALU0 9.40e-01
US-09-864-761-26966

Query Match          1.0%; Score 19; DB 10; Length 445;
Best Local Similarity 100.0%; Pred. No. 6;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 370 AGCTGTTCTTCTGATCTT 388
DB 286 AGCTGTTCTTCTGATCTT 268

RESULT 5
US-09-864-761-32700/c
; Sequence 32700, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecm1ca-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
```

US-09-864-761-32700
OTHER INFORMATION: EST HUMAN HIT: BE264998.1, EVALU0.000e+00

Query Match 1.0%; Score 19; DB 10; Length 561;
Best Local Similarity 100.0%; Pred. No. 6.2;

Qy 370 AGCTGTTCTTCTGAATCTT 388
Db 356 AGCTGTTCTTCTGAATCTT 388

RESULT 8
US-09-938-842A-186
; Sequence 186, Application US/09938842A
; Patent No. US20020160378A1

```

; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; TITLE OF INVENTION: SAME, AND METHODS OF USE
; FILE REFERENCE: SCRIPI300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 186
; LENGTH: 1707
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; US-09-938-842A-186

Query Match 1.0%; Score 19; DB 9; Length 1707;
Best Local Similarity 100.0%; Pred. No. 6.9;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1478 TTGATCATTTTGTATCGTA 1496
DB 673 TTGATCATTTTGTATCGTA 691

RESULT 9
US-09-938-842A-4668
; Sequence 4668, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; TITLE OF INVENTION: SAME, AND METHODS OF USE
; FILE REFERENCE: SCRIPI300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 4668
; LENGTH: 2000
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; US-09-938-842A-4668

Query Match 1.0%; Score 19; DB 9; Length 2000;
Best Local Similarity 100.0%; Pred. No. 7;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1395 ATTCTTTTGAAGAAAT 1413
DB 494 ATTCTTTTGAAGAAAT 512

RESULT 10
US-10-044-090-516/c
; Sequence 516, Application US/10044090
; Patent No. US20020137081A1
; GENERAL INFORMATION:
```

```

; APPLICANT: Olga Bandman
; TITLE OF INVENTION: GENES DIFFERENTIALLY EXPRESSED IN VASCULAR TISSUE ACTIVATION
; FILE REFERENCE: PA-0028 US
; CURRENT APPLICATION NUMBER: US/10/044,090
; CURRENT FILING DATE: 2002-01-09
; NUMBER OF SEQ ID NOS: 850
; SOFTWARE: PERL Program
; SEQ ID NO 516
; LENGTH: 3470
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20020137081A1 1994709CB1
; US-10-044-090-516

Query Match 1.0%; Score 19; DB 12; Length 3470;
Best Local Similarity 100.0%; Pred. No. 7.4;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1280 TTGGCAATCATCTTGCTCT 1298
DB 1712 TTGGCAATCATCTTGCTCT 1694

RESULT 11
US-09-878-574-2011
; Sequence 2011, Application US/09878574
; Patent No. US20020110548A1
; GENERAL INFORMATION:
; APPLICANT: Byrum, Joseph R.
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Thompson, Michael D.
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(15401)B
; CURRENT APPLICATION NUMBER: US/09/878,574
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 09/333,535
; PRIOR FILING DATE: 1999-06-14
; NUMBER OF SEQ ID NOS: 15775
; SEQ ID NO 2011
; LENGTH: 366
; TYPE: DNA
; ORGANISM: Glycine max
; OTHER INFORMATION: Clone ID: LIB3028-030-01-B1-C1
; US-09-878-574-2011

Query Match 1.0%; Score 18; DB 10; Length 366;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1785 TAGCAATATAAAATCTC 1802
DB 7 TAGCAATATAAAATCTC 24

RESULT 12
US-09-924-035A-470
; Sequence 470, Application US/09924035A
; Patent No. US20020142319A1
; GENERAL INFORMATION:
; APPLICANT: Grilach, Jrn
; TITLE OF INVENTION: Expressed Sequences of Arabidopsis
; FILE REFERENCE: 2011US
; CURRENT APPLICATION NUMBER: US/09/924,035A
; CURRENT FILING DATE: 2000-08-11
; PRIOR APPLICATION NUMBER: US 60/148,784
; PRIOR FILING DATE: 1999-08-13
; NUMBER OF SEQ ID NOS: 900
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 470
```



```

; LENGTH: 392
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-924-035A-470

Query Match      1.0%; Score 18; DB 10; Length 392;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 517 ATTTTATTGCTCAAG 534
      |||||||
Db 66 ATTTTATTGCTCAAG 83

RESULT 13
US-09-983-965-2819/c
; Sequence 2819, Application US/09983965
; Patent No. US20020137160A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION
; FILE REFERENCE: 37-21(10297)C
; CURRENT APPLICATION NUMBER: US/09/983,965
; CURRENT FILING DATE: 2001-10-26
; PRIOR APPLICATION NUMBER: US 09/465,231
; PRIOR FILING DATE: 1999-12-15
; PRIOR APPLICATION NUMBER: US 60/113,678
; PRIOR FILING DATE: 1998-12-17
; NUMBER OF SEQ ID NOS: 5912
; SEQ ID NO 2819
; LENGTH: 421
; TYPE: DNA
; ORGANISM: Bos taurus
; FEATURE:
; OTHER INFORMATION: Clone ID: 15-LIB3058-038-Q1-K1-D3
US-09-983-965-2819

Query Match      1.0%; Score 18; DB 10; Length 421;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 713 ACTAAATTTTCTTCATG 730
      |||||||
Db 268 ACTAAATTTTCTTCATG 251

RESULT 14
US-09-920-300A-17/c
; Sequence 17, Application US/09920300A
; Patent No. US20020136728A1
; GENERAL INFORMATION:
; APPLICANT: King, Gordon E.
; APPLICANT: Meagher, Madeleine Joy
; APPLICANT: Xu, Jiangchun
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121.547
; CURRENT APPLICATION NUMBER: US/09/920,300A
; CURRENT FILING DATE: 2001-07-31
; NUMBER OF SEQ ID NOS: 1789
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 459
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-920-300A-17

Query Match      1.0%; Score 18; DB 10; Length 459;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 713 ACTAAATTTTCTTCATG 730
      |||||||
Db 268 ACTAAATTTTCTTCATG 251
```

```

Best Local Similarity 100.0%; Pred. No. 20;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 887 TCCGTTTGGGAATGGAT 904
      |||||||
Db 36 TCCGTTTGGGAATGGAT 19

RESULT 15
US-10-033-528-17/c
; Sequence 17, Application US/10033528
; Patent No. US20020131971A1
; GENERAL INFORMATION:
; APPLICANT: King, Gordon E.
; APPLICANT: Meagher, Madeleine Joy
; APPLICANT: Xu, Jiangchun
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121.547C1
; CURRENT APPLICATION NUMBER: US/10/033,528
; CURRENT FILING DATE: 2001-12-26
; NUMBER OF SEQ ID NOS: 1896
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 459
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-033-528-17

Query Match      1.0%; Score 18; DB 12; Length 459;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 887 TCCGTTTGGGAATGGAT 904
      |||||||
Db 36 TCCGTTTGGGAATGGAT 19

RESULT 16
US-09-770-444-3/c
; Sequence 3, Application US/09770444
; Patent No. US20020023280A1
; GENERAL INFORMATION:
; APPLICANT: Goxlach, Jorn
; APPLICANT: An, Yong-Qiang
; APPLICANT: Hamilton, Carol M.
; APPLICANT: Price, Jennifer L.
; APPLICANT: Raines, Tracy M.
; APPLICANT: Yu, Yang
; APPLICANT: Rameaka, Joshua G.
; APPLICANT: Page, Amy
; APPLICANT: Matthew, Abraham V.
; APPLICANT: Ledford, Brooke L.
; APPLICANT: Woessner, Jeffrey P.
; APPLICANT: Haas, William David
; APPLICANT: Garcia, Carlos A.
; APPLICANT: Kriker, Maja
; APPLICANT: Slader, Ted
; APPLICANT: Davis, Keith R.
; APPLICANT: Allen, Keith
; APPLICANT: Hoffman, Neil
; APPLICANT: Hurban, Patrick
; TITLE OF INVENTION: Expressed Sequences of Arabidopsis
; TITLE OF INVENTION: thaliana
; FILE REFERENCE: 2027 (PARA-016PRV)
; CURRENT APPLICATION NUMBER: US/09/770,444
; CURRENT FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: 60/178,502
; PRIOR FILING DATE: 2000-01-27
; NUMBER OF SEQ ID NOS: 999
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
```

LENGTH: 478
TYPE: DNA
ORGANISM: Arabidopsis thaliana
US-09-770-444-3

Query Match 1.0%; Score 18; DB 10; Length 478;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 835 TCGAAGTGTCTTTCTTT 852
Db 97 TCGAAGTGTCTTTCTTT 80

RESULT 17
US-09-770-149-852
Sequence 852, Application US/09770149
Patent No. US20020059663A1
GENERAL INFORMATION:
APPLICANT: Goriach, Jorn
APPLICANT: An, Yong-Qiang
APPLICANT: Hamilton, Carol M.
APPLICANT: Price, Jennifer L.
APPLICANT: Raines, Tracy M.
APPLICANT: Yu, Yang
APPLICANT: Rameaka, Joshua G.
APPLICANT: Page, Amy
APPLICANT: Matthew, Abraham V.
APPLICANT: Ledford, Brooke L.
APPLICANT: Moesener, Jeffrey P.
APPLICANT: Haas, William David
APPLICANT: Garcia, Carlos A.
APPLICANT: Kicker, Maja
APPLICANT: Slader, Ted
APPLICANT: Davis, Keith R.
APPLICANT: Allen, Keith
APPLICANT: Hoffman, Neil
APPLICANT: Hurlan, Patrick
TITLE OF INVENTION: Expressed Sequences of Arabidopsis
FILE REFERENCE: 2024 (PARA-013PRV)
CURRENT APPLICATION NUMBER: US/09/770,149
CURRENT FILING DATE: 2001-01-26
PRIOR APPLICATION NUMBER: 60/178,506
PRIOR FILING DATE: 2000-01-27
NUMBER OF SEQ ID NOS: 999
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 852
LENGTH: 613
TYPE: DNA
ORGANISM: Arabidopsis thaliana
US-09-770-149-852

Query Match 1.0%; Score 18; DB 10; Length 613;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 517 ATTTTATGTCCTCAAG 534
Db 73 ATTTTATGTCCTCAAG 90

RESULT 18
US-09-764-847-1360
Sequence 1360, Application US/09764847
Patent No. US20020132767A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PC009
CURRENT APPLICATION NUMBER: US/09/764,847
CURRENT FILING DATE: 2001-01-17
Prior application data removed - consult PALM or file wrapper

NUMBER OF SEQ ID NOS: 2003
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1360
LENGTH: 713
TYPE: DNA
ORGANISM: Homo sapiens
US-09-764-847-1360

Query Match 1.0%; Score 18; DB 10; Length 713;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 617 GCACCTCTTTCTTCTTA 634
Db 254 GCACCTCTTTCTTCTTA 271

RESULT 19
US-09-938-842A-5222
Sequence 5222, Application US/09938842A
Patent No. US20020160378A1
GENERAL INFORMATION:
APPLICANT: Harper, Jeff
APPLICANT: Kreps, Joel
APPLICANT: Wang, Xun
APPLICANT: Zhu, Tong
TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
FILE REFERENCE: SCRIPI300-3
CURRENT APPLICATION NUMBER: US/09/938,842A
CURRENT FILING DATE: 2001-08-24
PRIOR APPLICATION NUMBER: US 60/227,866
PRIOR FILING DATE: 2000-08-24
PRIOR APPLICATION NUMBER: US 60/264,647
PRIOR FILING DATE: 2001-01-16
PRIOR APPLICATION NUMBER: US 60/300,111
PRIOR FILING DATE: 2001-06-22
NUMBER OF SEQ ID NOS: 5379
SEQ ID NO 5222
LENGTH: 755
TYPE: DNA
ORGANISM: Arabidopsis thaliana
US-09-938-842A-5222

Query Match 1.0%; Score 18; DB 9; Length 755;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1707 TTGCACCACTCTGTGT 1724
Db 367 TTGCACCACTCTGTGT 384

RESULT 20
US-09-887-576-715
Sequence 715, Application US/09887576
Patent No. US20020144047A1
GENERAL INFORMATION:
APPLICANT: Budworth, P.
APPLICANT: Brown, D.
APPLICANT: Chang, H.
APPLICANT: Zhu, T.
APPLICANT: Han, B.
APPLICANT: Wang, X.
APPLICANT: Cooper, Bret
TITLE OF INVENTION: Promoters for regulation of plant expression
FILE REFERENCE: 1360.001US1
CURRENT APPLICATION NUMBER: US/09/887,576
CURRENT FILING DATE: 2001-06-25
PRIOR APPLICATION NUMBER: US 60/213,848
PRIOR FILING DATE: 2000-06-23
PRIOR APPLICATION NUMBER: US 60/214,087
PRIOR FILING DATE: 2000-06-23


```

; APPLICANT: Chen, Mansheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecmlca-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annonmax Sequence Listing Engine vers. 1.1
; SEQ ID NO 26319
; LENGTH: 178
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC021498.4
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.4
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3.2
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.9
; OTHER INFORMATION: EST HUMAN HIT: AL041184.2, EVALUATE 9.60e+00
; OTHER INFORMATION: SWISSPROT HIT: O59295, EVALUATE 3.90e+00
; OTHER INFORMATION: NT HIT: X52046.1, EVALUATE 1.60e+00
; US-09-864-761-26319

```

```

Query Match 0.9%; Score 17; DB 10; Length 178;
Best Local Similarity 100.0%; Pred. No. 63;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 658 TGCTTGAACATGATTT 674
|||
Db 98 TGCTTGAACATGATTT 114

```

```

RESULT 26
US-10-046-935-1387

```

```

; Sequence 1387, Application US/10046935
; Patent No. US2002015601A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yugu
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Secrist, Heather
; APPLICANT: Wang, Aijun
; APPLICANT: Stolk, John A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.527C1
; CURRENT APPLICATION NUMBER: US/10/046,935
; CURRENT FILING DATE: 2002-01-15
; NUMBER OF SEQ ID NOS: 2239
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1387
; LENGTH: 293
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-10-046-935-1387

```

```

Query Match 0.9%; Score 17; DB 9; Length 293;
Best Local Similarity 100.0%; Pred. No. 66;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1609 TTGTTAATGCTTTTGT 1625
|||
Db 148 TTGTTAATGCTTTTGT 164

```

```

RESULT 27
US-09-878-178-1387
; Sequence 1387, Application US/09878178
; Patent No. US20020177552A1
; GENERAL INFORMATION:
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yugu
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.527
; CURRENT APPLICATION NUMBER: US/09/878,178
; CURRENT FILING DATE: 2001-06-08
; NUMBER OF SEQ ID NOS: 2237
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1387
; LENGTH: 293
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-878-178-1387

```

```

Query Match 0.9%; Score 17; DB 9; Length 293;
Best Local Similarity 100.0%; Pred. No. 66;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1609 TTGTTAATGCTTTTGT 1625
|||
Db 148 TTGTTAATGCTTTTGT 164

```

```

RESULT 28
US-09-962-436-452
; Sequence 452, Application US/09962436
; Patent No. US20020081301A1
; GENERAL INFORMATION:
; APPLICANT: Soppet, Daniel
; TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Signa
; FILE REFERENCE: 689290-75
; CURRENT APPLICATION NUMBER: US/09/962,436
; CURRENT FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US/60/235,082
; PRIOR FILING DATE: 2000-09-25

```

```
; PRIOR APPLICATION NUMBER: US/60/234,924
; PRIOR FILING DATE: 2000-09-25
; NUMBER OF SEQ ID NOS: 568
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 452
; LENGTH: 324
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-962-436-452

Query Match          0.9%; Score 17; DB 10; Length 324;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 191 TCAGATTAAATCTGG 207
    |||||
Db 153 TCAGATTAAATCTGG 169

RESULT 29
US-09-867-701-6153
; Sequence 6153, Application US/09867701
; Patent No. US20020132237A1
; GENERAL INFORMATION:
; APPLICANT: Aglate, Paul A.
; APPLICANT: Jones, Robert
; APPLICANT: Harlocker, Susan L.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.497
; CURRENT APPLICATION NUMBER: US/09/867,701
; CURRENT FILING DATE: 2001-05-29
; NUMBER OF SEQ ID NOS: 10912
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6153
; LENGTH: 403
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-867-701-6153

Query Match          0.9%; Score 17; DB 10; Length 403;
Best Local Similarity 100.0%; Pred. No. 69;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1783 TGTAGCAATATAAAAT 1799
    |||||
Db 67 TGTAGCAATATAAAAT 83

RESULT 30
US-09-983-965-3885
; Sequence 3885, Application US/09983965
; Patent No. US20020137160A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 37-21(10297)C
; CURRENT APPLICATION NUMBER: US/09/983,965
; CURRENT FILING DATE: 2001-10-26
; PRIOR FILING DATE: 1999-12-15
; PRIOR APPLICATION NUMBER: US 60/113,678
; PRIOR FILING DATE: 1998-12-17
; NUMBER OF SEQ ID NOS: 5912
; SEQ ID NO 3885
; LENGTH: 434
; TYPE: DNA
; ORGANISM: Bos taurus
; FEATURE:
```

```
; OTHER INFORMATION: Clone ID: 50-LIB3058-007-Q1-K1-E6
US-09-983-965-3885

Query Match          0.9%; Score 17; DB 10; Length 434;
Best Local Similarity 100.0%; Pred. No. 69;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 843 TTTTCTTTTAGTTC 859
    |||||
Db 366 TTTTCTTTTAGTTC 382

RESULT 31
US-09-867-701-5590/c
; Sequence 5590, Application US/09867701
; Patent No. US20020132237A1
; GENERAL INFORMATION:
; APPLICANT: Aglate, Paul A.
; APPLICANT: Jones, Robert
; APPLICANT: Harlocker, Susan L.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.497
; CURRENT APPLICATION NUMBER: US/09/867,701
; CURRENT FILING DATE: 2001-05-29
; NUMBER OF SEQ ID NOS: 10912
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5590
; LENGTH: 478
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-867-701-5590

Query Match          0.9%; Score 17; DB 10; Length 478;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1459 GTATTCAATAACAATA 1475
    |||||
Db 423 GTATTCAATAACAATA 407

RESULT 32
US-09-998-598-1850/c
; Sequence 1850, Application US/09998598
; Patent No. US20020150922A1
; GENERAL INFORMATION:
; APPLICANT: Stolk, John A.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Meagher, Madelein Joy
; APPLICANT: Chenault, Ruth A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.561
; CURRENT APPLICATION NUMBER: US/09/998,598
; CURRENT FILING DATE: 2001-11-16
; NUMBER OF SEQ ID NOS: 2606
; SOFTWARE: Corixa Invention Disclosure Database
; SEQ ID NO 1850
; LENGTH: 478
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-998-598-1850

Query Match          0.9%; Score 17; DB 10; Length 478;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 426 CACTCACCTTGTGAGGA 442
    |||||
Db 433 CACTCACCTTGTGAGGA 417
```

RESULT 33
US-09-864-761-10630/C
Sequence 10630, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aecm1ca-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
SEQ ID NO 10630
LENGTH: 494
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC005392.1
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.91
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.9
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.94
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.77
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.99
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.8
US-09-864-761-10630

Query Match 0.9%; Score 17; DB 10; Length 494;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1319 TGGCAATATGCTTTCAG 1335
|||||
Db 263 TGGCAATATGCTTTCAG 247

RESULT 34
US-09-734-569-67
Sequence 67, Application US/09734569
Patent No. US20020064816A1
GENERAL INFORMATION:
APPLICANT: Lerchl, Jens
APPLICANT: Renz, Andreas
APPLICANT: Ehrhardt, Thomas
APPLICANT: Reindel, Andreas
APPLICANT: Cippus, Petra
APPLICANT: Bischoff, Friedrich
APPLICANT: Freund, Markus
APPLICANT: Duwendt, Elke
APPLICANT: Schmidt, Ralf-Michael
APPLICANT: Reski, Ralf
TITLE OF INVENTION: Moss genes from Physcomitrella patens encoding proteins involve
FILE REFERENCE: BASF-NAE-1332-99-US
CURRENT APPLICATION NUMBER: US/09/734,569
CURRENT FILING DATE: 2001-05-24
PRIOR APPLICATION NUMBER: US 60/171,101
PRIOR FILING DATE: 1999-12-16
NUMBER OF SEQ ID NOS: 181
SOFTWARE: Patentin Ver. 2.1/Wordperfect 6.1
SEQ ID NO 67
LENGTH: 496
TYPE: DNA
ORGANISM: Physcomitrella patens
FEATURE:
NAME/KEY: CDS
LOCATION: (1)...(297)
OTHER INFORMATION: 94_ck14_h11fwd
US-09-734-569-67

Query Match 0.9%; Score 17; DB 10; Length 496;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1572 GCCGACGGTATGACCA 1588
|||||
Db 146 GCCGACGGTATGACCA 162

RESULT 35
US-09-864-761-7444
Sequence 7444, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aecm1ca-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667

```
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 7444
; LENGTH: 539
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC021498.4
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.4
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3.2
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.9
; US-09-864-761-7444
```

```
Query Match 0.9%; Score 17; DB 10; Length 539;
Best Local Similarity 100.0%; Pred. No. 71;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 658 TGCTTTGACATTGATT 674
| | | | | | | | | |
Db 173 TGCTTTGACATTGATT 189
```

```
RESULT 36
US-09-734-569-65
; Sequence 65, Application US/09734569
; Patent No. US20020064816A1
; GENERAL INFORMATION:
; APPLICANT: Lerchl, Jens
; APPLICANT: Renz, Andreas
; APPLICANT: Ehrhardt, Thomas
; APPLICANT: Reindl, Andreas
; APPLICANT: Cirpus, Petra
; APPLICANT: Bischoff, Friedrich
; APPLICANT: Frank, Markus
; APPLICANT: Freund, Annette
; APPLICANT: Duwenig, Elke
; APPLICANT: Schmidt, Ralf-Michael
; APPLICANT: Reiki, Ralf
; TITLE OF INVENTION: Moss genes from Physcomitrella patens encoding proteins involved
; FILE REFERENCE: BASF-NAE-1332-99-US
; CURRENT APPLICATION NUMBER: US/09/734,569
; CURRENT FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: US 60/171,101
; PRIOR FILING DATE: 1999-12-16
; NUMBER OF SEQ ID NOS: 181
```

```
; SOFTWARE: PatentIn Ver. 2.1/WordPerfect 6.1
; SEQ ID NO 65
; LENGTH: 548
; TYPE: DNA
; ORGANISM: Physcomitrella patens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(375)
; OTHER INFORMATION: 71_ck18_d06fwd
US-09-734-569-65
```

```
Query Match 0.9%; Score 17; DB 10; Length 548;
Best Local Similarity 100.0%; Pred. No. 71;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 1572 GCCGCACGGTAATGACA 1588
| | | | | | | | | |
Db 224 GCCGCACGGTAATGACA 240
```

```
RESULT 37
US-09-969-347-157
; Sequence 157, Application US/09969347
; Patent No. US20020115085A1
; GENERAL INFORMATION:
; APPLICANT: Ebner, Reinhard
; TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Signatu
; TITLE OF INVENTION: Sets
; FILE REFERENCE: 689290-69
; CURRENT APPLICATION NUMBER: US/09/969,347
; CURRENT FILING DATE: 2001-10-02
; PRIOR APPLICATION NUMBER: US/60/237,598
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: US/60/237,604
; PRIOR FILING DATE: 2000-10-03
; NUMBER OF SEQ ID NOS: 318
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 157
; LENGTH: 599
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: n=a,t,g or c
US-09-969-347-157
```

```
Query Match 0.9%; Score 17; DB 10; Length 599;
Best Local Similarity 100.0%; Pred. No. 71;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 191 TCAGATTAAATCTGG 207
| | | | | | | | | |
Db 157 TCAGATTAAATCTGG 173
```

```
RESULT 38
US-09-999-162-9
; Sequence 9, Application US/09999162
; Patent No. US20020137171A1
; GENERAL INFORMATION:
; APPLICANT: Short, Jay
; APPLICANT: Mathur, Eric J.
; APPLICANT: Baumann, Markus
; APPLICANT: Bornscheuer, Uwe T.
; TITLE OF INVENTION: Hydrolase Enzymes and Their Use in Kinetic Resolution
; FILE REFERENCE: DIV-011 US/112766.126
; CURRENT APPLICATION NUMBER: US/09/999,162
; CURRENT FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER: US 60/250,976
; PRIOR FILING DATE: 2000-12-01
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 9
```

```

; LENGTH: 933
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: polynucleotide BD073 encoding polypeptide with
; OTHER INFORMATION: hydrolase activity
US-09-999-162-9

Query Match          0.9%; Score 17; DB 10; Length 933;
Best Local Similarity 100.0%; Pred. No. 75;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1232 CTATCGAGAAATGCG 1248
Db 587 CTATCGAGAAATGCG 603

RESULT 39
US-09-734-569-161
; Sequence 161, Application US/09734569
; Patent No. US20020064816A1
; GENERAL INFORMATION:
; APPLICANT: Lerchl, Jens
; APPLICANT: Renz, Andreas
; APPLICANT: Ehrhardt, Thomas
; APPLICANT: Reinol, Andreas
; APPLICANT: Cirus, Petra
; APPLICANT: Bischoff, Friedrich
; APPLICANT: Frank, Markus
; APPLICANT: Freund, Annette
; APPLICANT: Duwenig, Elke
; APPLICANT: Schmidt, Ralf-Michael
; APPLICANT: Reski, Ralf
; TITLE OF INVENTION: Moss genes from Physcomitrella patens encoding proteins involved
; TITLE OF INVENTION: in the synthesis of carbohydrates
; FILE REFERENCE: BASF-NAE-1332-99-US
; CURRENT APPLICATION NUMBER: US/09/734,569
; CURRENT FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: US 60/171,101
; PRIOR FILING DATE: 1999-12-16
; NUMBER OF SEQ ID NOS: 181
; SOFTWARE: Patentin Ver. 2.1/WordPerfect 6.1
; SEQ ID NO 161
; LENGTH: 979
; TYPE: DNA
; ORGANISM: Physcomitrella patens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(798)
; OTHER INFORMATION: c_dp001074086r
US-09-734-569-161

Query Match          0.9%; Score 17; DB 10; Length 979;
Best Local Similarity 100.0%; Pred. No. 75;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1572 GCCGCACGTAATGACA 1588
Db 647 GCCGCACGTAATGACA 663

RESULT 40
US-09-938-842A-5070
; Sequence 5070, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; TITLE OF INVENTION: SAME, AND METHODS OF USE
; FILE REFERENCE: SRIPI300-3
```

```

; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 5070
; LENGTH: 1121
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-5070

Query Match          0.9%; Score 17; DB 9; Length 1121;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1228 CAAACTATCGAAGAAA 1244
Db 988 CAAACTATCGAAGAAA 1004

RESULT 41
US-09-815-242-9546/c
; Sequence 9546, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zykkind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9546
; LENGTH: 1551
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)....(1551)
US-09-815-242-9546

Query Match          0.9%; Score 17; DB 10; Length 1551;
Best Local Similarity 100.0%; Pred. No. 78;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1762 ACCGTAACCTATATCT 1778
Db 1762 ACCGTAACCTATATCT 1778
```


Db 1239 ACCGTTAAACTTAATCT 1223

RESULT 42

US-09-939-980-116/c
; Sequence 116, Application US/09939980
; Patent No. US2002008234A1

GENERAL INFORMATION:

APPLICANT: Black, Michael
Burnham, Martin
Hodgson, John
Knowles, David
Lonetto, Michael
Nicholas, Richard
Pratt, Julie
Reichard, Richard
Rosenberg, Martin
Ward, Judith

TITLE OF INVENTION: No. US2002008234A1el Prokaryotic Polynucleotides,
Polypeptides and Their Uses

NUMBER OF SEQUENCES: 534

CORRESPONDENCE ADDRESS:

ADDRESSEE: Smithkline Beecham Corporation
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA

COUNTRY: USA

ZIP: 19406-0939

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSEQ for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/939,980

FILING DATE: 27-Aug-2001

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/936,165

FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Gimmi, Edward R

REGISTRATION NUMBER: 38,891

REFERENCE/DOCKET NUMBER: P50549

TELECOMMUNICATION INFORMATION:

TELEPHONE: 610-270-4478

TELEFAX: 610-270-5090

TELEX: <Unknown>

INFORMATION FOR SEQ ID NO: 116:

SEQUENCE CHARACTERISTICS:

LENGTH: 1789 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: Genomic DNA

SEQUENCE DESCRIPTION: SEQ ID NO: 116:

US-09-939-980-116

Query Match

Best Local Similarity 100.0%; Score 17; DB 10; Length 1789;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 189 CCTCAGATTTAAATCT 205

Db 94 CCTCAGATTTAAATCT 78

RESULT 43

US-09-887-576-173

; Sequence 173, Application US/09887576

; Patent No. US20020144047A1

; GENERAL INFORMATION:

; APPLICANT: Budworth, P.

; APPLICANT: Brown, D.
; APPLICANT: Chang, H.
; APPLICANT: Zhu, T.
; APPLICANT: Han, B.
; APPLICANT: Wang, X.
; APPLICANT: Cooper, Bret

; TITLE OF INVENTION: Promoters for regulation of plant expression

; FILE REFERENCE: 1360.001US1

; CURRENT APPLICATION NUMBER: US/09/887,576

; CURRENT FILING DATE: 2001-06-25

; PRIOR APPLICATION NUMBER: US 60/213,848

; PRIOR FILING DATE: 2000-06-23

; PRIOR APPLICATION NUMBER: US 60/214,087

; PRIOR FILING DATE: 2000-06-23

; PRIOR APPLICATION NUMBER: US 60/258,692

; PRIOR FILING DATE: 2000-12-29

; NUMBER OF SEQ ID NOS: 875

; SOFTWARE: FastSEQ for Windows Version 4.0

; SEQ ID NO 173

; LENGTH: 2004

; TYPE: DNA

; ORGANISM: Arabidopsis thaliana

US-09-887-576-173

Query Match

Best Local Similarity 100.0%; Score 17; DB 10; Length 2004;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 618 CACTTCTTTCTTCCTA 634

Db 110 CACTTCTTTCTTCCTA 126

RESULT 44

US-09-950-788-1/c

; Sequence 1, Application US/09950788

; Patent No. US2002013232A1

; GENERAL INFORMATION:

APPLICANT: MOECKEL, Bettina

APPLICANT: HANS, Stephan

APPLICANT: SCHISCHKA, Natalie

APPLICANT: PFEFFERLE, Walter

; TITLE OF INVENTION: NUCLEOTIDE SEQUENCES WHICH CODE FOR THE PT51 PROTEIN

; FILE REFERENCE: 213544US0

; CURRENT APPLICATION NUMBER: US/09/950,788

; CURRENT FILING DATE: 2001-09-13

; PRIOR APPLICATION NUMBER: DE10045496.8

; PRIOR FILING DATE: 2000-09-13

; NUMBER OF SEQ ID NOS: 9

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 1

; LENGTH: 2005

; TYPE: DNA

; ORGANISM: Corynebacterium glutamicum

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (154)..(1857)

; OTHER INFORMATION:

US-09-950-788-1

Query Match

Best Local Similarity 100.0%; Score 17; DB 10; Length 2005;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1027 TATTCAAACATTTGGAT 1043

Db 91 TATTCAAACATTTGGAT 75

RESULT 45

US-09-950-788-3/c

; Sequence 3, Application US/09950788

; Patent No. US2002013232A1

```
; GENERAL INFORMATION:
; APPLICANT: MOECKEL, Bettina
; APPLICANT: HANS, Stephan
; APPLICANT: SCHISCHKA, Natalie
; APPLICANT: PEEFERLE, Walter
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCES WHICH CODE FOR THE PTSI PROTEIN
; FILE REFERENCE: 213544US0
; CURRENT APPLICATION NUMBER: US/09/950,788
; CURRENT FILING DATE: 2001-09-13
; PRIOR APPLICATION NUMBER: DE10045496.8
; PRIOR FILING DATE: 2000-09-13
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 2005
; TYPE: DNA
; ORGANISM: Corynebacterium glutamicum
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (154)..(1857)
; OTHER INFORMATION:
US-09-950-788-3
```

```
Query Match          0.9%; Score 17; DB 10; Length 2005;
Best Local Similarity 100.0%; Pred.No.80;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Oy 1027 TATTCAAAACATTGGAT 1043
      |||||
Db 91 TATTCAAAACATTGGAT 75
```

Search completed: December 17, 2002, 11:36:00
Job time : 148 secs

```

MEDLINE 99206606
PUBMED 10192398
REFERENCE 2 (bases 1 to 14134)
AUTHORS Kalman,S., Mitchell,W., Marathe,R., Lammel,C., Fan,J., Olinger,L.,
Grimwood,J., Davis,R.W. and Stephens,R.S.
TITLE Direct Submission
JOURNAL Submitted (01-DEC-1998) Program in Infectious Diseases, University
of California, 235 Earl Warren Hall, Berkeley, CA 94720, USA
FEATURES
source
Location/Qualifiers
1. .14134
/organism="Chlamydothila pneumoniae CWL029"
/strain="CWL029"
/db_xref="taxon:115713"
complement (158..1924)
/gene="recJ"
/gene="recJ"
note="CpN0563"
complement (158..1924)
/gene="recJ"
/codon_start=1
/transl_table=11
/product="ssDNA Exonuclease"
/protein_id="AAD18703.1"
/db_xref="GI:4376856"
/translation="MTNSDNASAGLLWAHKEDPAFLGMIKEFHLPTVAQIFISR
GFOTIOETHKLYSHLSLVDPLGFLDMASKAVERILLARDKHEHVMYIGSDVDGWTG
VALLVEFLRDIDVHVSYPFLGAILKQHGSETSLAKLKEGITILLITVDGCGITAGKEV
SOTRQGDIVITDHHMETGKIPHCVTAINPKLRDHYTFNRELTVGVGFAKLGAVLN
ALISRLNPKSQGLKLLDLVTLTITDVGVLGKGNRMVYRGIKEIKARAGPLNK
LCAICGVSEVTSTDIIVLKIAPKLSLGRLLDDPAKGVLELLITQDDERVALIMELN
INRQRTEAVFOVDQVILNSPEILKQAAIIVLSSTAMHARVPIISARLAKTYNKP
VVIIAIORGSGKARTIGSPFLGLVKKCSLLSYGGHDFAAGVIMKEDKVDKFK
KRVHLVNSLKKGDTPLHLEIDAYDAFDIYDILLASNELEFEPFGKGNLMPFIYSKVR
QVRYPVLPFHNKLIYLSQKERNLEGVAGFLGRHADALKAWSHYPLEIATYTPRUSQTS
GSGVHLHVRDFRISSEPRFSD"
complement (2010..6218)
/gene="secD/secF"
/gene="secD/secF"
note="CpN0564"
complement (2010..6218)
/gene="secD/secF"
/codon_start=1
/transl_table=11
/product="Protein Export Proteins SecD/SecF (fusion)"
/protein_id="RAD18704.1"
/db_xref="GI:4376857"
/translation="MKQVKKNFAIIICVFALALYYVLPTCLYVAKPLKKIDGNEAB
HIKFTFOAQOVRKDVIPRVSAILSSHLRGHIQOHPAIDIVSVRFRGEDAEFI
GNLHGEPNVPIKSARLHVGVYSREHDDHVIQVASSINTSLVESDFSVYSSENEOE
MASSILQRYVACTPPOKDCSCSYPIWETAPKQLQYAKNLSGFEVPSRLSAP
COQFSSNODRLAFLRSLSSNDAAIDVEDQKLKSVETLSOTACIRSLDCPYIEG
LRLDSESLSFSSIEYCPKERRKIFLTHSLDLAORTSLSKERLDFDSRLAVKQKL
SKNLTUVQEDYNGSFQWMDKDTQKIILOGERLLQGIASHLTALTUHRPAASSCDL
IPENPFVCRPQRESEAFGCYIFSTDCRKFSGSVYLLKGRISIVAKYQQGKGKE
LQSFEDLQNLNYNCSFHTTEAISWTLPQDQVLEIRHPLQQLDVMGEGFVKEGECFL
EVKIDQDLATVNOIEKNQSDLVWHEQYRHAQKMDQLRSLAPIYONLFLENMK
LNMKRSRGENTLRIGIDFVGGROLLSFKDHQKQLTDKEDILKVSDELARLNKLG
VSEILRGENGYIHLSVPGSSTISSEILGTSKSHVNNRFSYSASRYEVRFLD
LYMFTSQAGKTSVPEINTFASLFNEEDVPPSVHEAITKLKSEGLAFSPSGCETPS
TDLDTFFSMIAKDAQEQAQNPVIFVRNALDGLASLKDIREFAAGGYYLVNFSVD
TSPTKMAKSLSTESFHTWTSAYCOEGISGTANGQYSANRGMWAVIDGVMVSPIL
NVLKNAHVSQKFTHREVSKLASDLKSGAMSFPVSEVSEETISDLGKCKCTGGIIS
ACCGALLVMSVYVRFGVATSCAVLNLILLIWAALOYLDAPLTSLGLAGIVLWG
MAVDANVIFERIREPRLLSQGLKSKVEGYTKAFCAIPDSNLTVLASALLPFLDTG
PIKGFALLTILIGIFSSMFTALFMTFFPMLWNKQTQTLHMMNKFPVGIKHDFLRGK
KLMVGSVFLGLGVALGAWNSVLGMDFKGGAFTFNPKHEGISDVAQMRGVVHK
LQEAGLSKDRFRIQTFGSEKIKYFSDKALSYTKADTSLSPKINDHELAVGLLSE
TGLDFTSTLTNETONFWSKSKMRYOATIGLIGALAIILLYVSLRFWEQYAFS
AVCALIHDLTAVLCAFLTAHFELKXIQLDQAGLMTVLGVSNNTLIIEDRIREDR
QANLTPHVLVNDALQKTFSTVMTIATISVLMLLFIQSSSVNFNFIWITGILL
GNTLSSTYAPPLLLFWVRKNRSK"
complement (6398..7498)
/gene="CpN0565"
complement (6398..7498)
/gene="CpN0565"

```

```

/codon_start=1
/transl_table=11
/product="Ct449 hypothetical protein"
/protein_id="AAD18705.1"
/db_xref="GI:4376858"
/translation="MKFPFRTISDLIPTOMVIMRGGNVYVPAQNLPKKILGGVL
ACFGALALCAAPAAAGVCTOTIPPCIGLMTLGLVGLGFAVLOYSKGMSFEPLPRETK
VPEKPNWLGCLSLQSKWKIRPGCYHYHPCQVEICEGSEIYTKIFOKKSDRNTSI
FLIQEMDQIALRGQIEKSLSRKTFATDPSVVSLLSEIQREEQQLDLPKVISWSSD
QASDRTHPKSAIYVNISDAAPQGRCYDAYTKAFFTVLDQIGDPNIVKGTIYVLT
PILGVDPDALKEEQENKLLSQAAFLYSAEQVAKMRREKQDSIRIKFIFTDPTSPTS
LYFSPHSSTPHSVTPISLSGFVGEQSYTFA"
8009..8761
/gene="yaeS"
/note="CpN0566"
8009..8761
/gene="yaeS"
/codon_start=1
/transl_table=11
/product="yaeS family hypothetical protein"
/protein_id="AAD18706.1"
/db_xref="GI:4376859"
/translation="MSLATNNAESKFPSSLQRLPNHVAIIMDGNRRWYKKBRECCGHTH
TSGHYGAKVLPNILNVLVDLGIKVLTVTFSTENFRPKKEIOEINFIYTDLDKQL
PYLMENEICLCIGDLSKLPKGIQKINHVSMTASFSLVLELVAVGKQELVRAF
KXLHVDILNKKISSDLSLSSGLDTSGLDTPDLLIRTGEMRVSNFLMQIAYTE
LYITDVLDPDFTPDQLFEAINVYQKRSRGGK"
8762..9688
/gene="cdsA"
/note="CpN0567"
8762..9688
/gene="cdsA"
/codon_start=1
/transl_table=11
/product="Phosphatidate Cytidylyltransferase"
/protein_id="AAD18707.1"
/db_xref="GI:4376860"
/translation="MLNSNKFSTKTCAYGDLQFQVVVSHLVLTFLVLLYSLLPPLTS
FALGPITATCGAVGTVEYSMAKAMKHPLSTFSAIGSFLFALSFLSRWGHSLPGF
FDALPWTLLVVMVYSIFRKRSTIGALQSGVTLFSLYGLVPIRLHLVLYSFYHT
QESYLIQWASFLIATTKGADIFGVFGKAFGNKKIAPQISPNKTVGVFVAGCLGATL
ISPIFFLOIPTRFASYFPMFPAIILPLGLALGITGFFGDIIESIKRDAHLKNSKLKA
VSGMLDTLQSLLSLTPIALVFLLLITQSKFEG"
9685..10335
/gene="cmk"
/note="CpN0568"
9685..10335
/gene="cmk"
/codon_start=1
/transl_table=11
/product="CMP kinase"
/protein_id="AAD18708.1"
/db_xref="GI:4376861"
/translation="MITIDGSGTGKSTAKALADHLHFNVCNTGMVYRTLAYARLQ
SPWATPLPKLEEPPESTFATQGLSEFFNGHLLTSELTTQEVANAASELSQLEPV
RPMQOLRYAQLGNCVFEGRMDQKVPFNADLKIFLTSSPEVRAQRRLQDLPEGLT
SPQLQAEVLKDAADAQRAHDLPLVPENGVIDSSDLTIROVLEKILALLFRNEL"
10332..10970
/gene="plsC"
/note="CpN0569"
10332..10970
/gene="plsC"
/codon_start=1
/transl_table=11
/product="Glycerol-3-P Acyltransferase"
/protein_id="AAD18709.1"
/db_xref="GI:4376862"
/translation="MIFRICKFTTWAFSLFYKLKYGVKNFIKGPALIVNHNHSL
DPIALHMCVHECIYHLARSLFIPLWQKQWCFVRQDEGNSAFKIASRLFNKRRK
LVITYPEGARSPDQQLQKGVKIGMAAKSRVPIIPVYIRGTFEAFNRHOKIPIHWKTI
TCVFGTMYFDIIQNPEIKNKETYQIITNOTWNKIAELKAWYESCKGDVP"
10975..12660
/gene="argS"

```

THIS PAGE BLANK (USPTO)

CDS	10975..12660	/note="Cp00570"
	/gene="argS"	/protein_id="AAD18710.1"
	/codon_start=1	/db_xref="GI:4376863"
	/transl_table=11	/translation="MSTILSLTSLVICSQAIKAPFNADGAFINFTSPVFNQOHLHFQK
	/product="arginyl tRNA Transferase"	DAMKLAIVKKAPRAIAEAVELPOEPFSLIETAGFVINFSPVFNQOHLHFQK
	/protein_id="AAD18710.1"	ALKLGFOVSOPKKIILDPSSPNIAKDMHVGHLRSTIIGDSLARIFSYVGHVDVLRNLHI
	/db_xref="GI:4376863"	GDWGTAFGLMITYLOENPCDYSLELDTLSLYKKAIVCFYNDEEPKRSQOQNVLALQAK
	/transl_table=11	DPAQIAIWEKICETSEKPAQKIYDILDIVBKRGESFYNPFELPEIIEOLEKGLLTVS
	/product="arginyl tRNA Transferase"	NDKACVTHEAFSIFPMQSGDGGNYATVDLNAIRYIEEDHAKIIVITDGLGSLHF
	/protein_id="AAD18710.1"	QLLEBATAAGLYOIPGFSHVGFGLVDPOQKLGRTSGENVKRLRELLDTAIEKAEFA
	/db_xref="GI:4376863"	LREHRPELTDRAIPERAVIGNAIKIDLSHRTSDYVFPFKMLREGNTAMPLLY
	/transl_table=11	LYVLTQIKRRLGFSQLSLEGPPEIKQEPAEELLALTILRFEALRESTIKELCPHPLTD
	/product="arginyl tRNA Transferase"	AYVNLTKFNGFRFSDISQSFYAKSRFLFCALAEQVILATGHHLLGLTLERL"
	/protein_id="AAD18710.1"	/complement(12614..13990)
	/db_xref="GI:4376863"	/gene="murA"
	/transl_table=11	/note="Cp00571"
	/product="UDP-N-Acetylglucosamine Transferase"	/complement(12614..13990)
	/protein_id="AAD18710.1"	/db_xref="GI:4376864"
	/db_xref="GI:4376864"	/translation="MQIARFCCGRLNGEVKSGAKNAATKLLVASLLSDOKCTLRNV
	/transl_table=11	PDIGDVLTELUCKSLGARHVSMDKETEVELEIYTPICQTRVPTFSNNVRIFILLGA
	/product="UDP-N-Acetylglucosamine Transferase"	LILGRCEGYVYPTGQDGAIGERTLNFHFGLKQGVQISSDSSGYAKAPRGLKNGIY
	/protein_id="AAD18710.1"	HLPSYGVATENLIIAAIHAKGRTVIKNVALEIIDLVLFLQKAGADITDGLKNDIT
	/db_xref="GI:4376864"	IFGTGSLGVDFHTIIPDKIEAASPGMAVVSQGRVVRNAKQELIIPFLKMLRSIGGG
	/transl_table=11	FLVSSGSEIVFOERPLVGVLETDVHGFCLTDQOPAVLLSQAGSSVHETVHEN
	/product="UDP-N-Acetylglucosamine Transferase"	RLGYDLHGLQHMGAECQLFHQCLSTKACRYAIGNPHPSAVIHGATPLWASHLVIPDLRA
	/protein_id="AAD18710.1"	GPAYVMAALIAEGGSIIEHTLLDRGYTNWVKLSLUGAKIQIFDMEOEELTTSFKS
	/db_xref="GI:4376864"	LALRDASL"
BASE COUNT	4180 a 2981 c 2805 g 4168 t	
ORIGIN		
Query Match	98.1%; Score 1829.4; DB 1; Length 14134;	
Best Local Similarity	99.8%; Pred. No. 0;	
Matches 1863; Conservative	0; Mismatches 1; Indels 3; Gaps 3;	
Oy	1	ATGGACTTCGCGATATTGTTCAGGAGGGGATCAGCGGCACCTGCTAATGACCAATATTCCTGC 60
Db	3771	ATGGACTTCGCGATATTGTTCAGGAGGGGATCAGCGGCACCTGCTAATGACCAATATTCCTGC 3712
Oy	61	AAACCGTGAGTCGCTAGTGGTATGACGGTTATATGGTTCAGCAGCCCTATTTT 120
Db	3711	AAACCGTGAGTCGCTAGTGGTATGACGGTTATATGGTTCAGCAGCCCTATTTT 3652
Oy	121	AAACGTGCCATTGAAAAATCATGCGCAGTGCTCAGGGAAATTTACCCACGTCAGGTGAG 180
Db	3651	AAACGTGCCATTGAAAAATCATGCGCAGTGCTCAGGGAAATTTACCCACGTCAGGTGAG 3592
Oy	181	CAAACTCGCCTCAGATTAAAACTCGAGCGATGCTCTTTTGTTCCTCCGAGGTTCTCAGTGA 240
Db	3591	CAAACTCGCCTCAGATTAAAACTCGAGCGATGCTCTTTTGTTCCTCCGAGGTTCTCAGTGA 3532
Oy	241	AGAGACGATCTCTTCTGATCTTGGGAAAAAACAATGTACACAAGGCATTTCTCAGCATG 300
Db	3531	AGAGACGATCTCTTCTGATCTTGGGAAAAAACAATGTACACAAGGCATTTCTCAGCATG 3472
Oy	301	CTGTGGCTTGGCAATGCTTATTTTGTAGAGCGGTATATATAGATTGAGGCGGTCTAT 360
Db	3471	CTGTGGCTTGGCAATGCTTATTTTGTAGAGCGGTATATATAGATTGAGGCGGTCTAT 3412
Oy	361	CGCTTCGGGAGCTCTTCTTGATCTTTTGTCTTATCTGGGACGCTCTACAGTATTTTGA 420
Db	3411	CGCTTCGGGAGCTCTTCTTGATCTTTTGTCTTATCTGGGACGCTCTACAGTATTTTGA 3352
Oy	421	TGCGCCACTCACCTTGTCCAGACTCGCTGGGATTTGTTCTTCTGATGGGATGCGCGTAGA 480

[illegible]

THIS PAGE BLANK (USPTO)

```

Db 2271 TCGTGAAGATCCCAAGCAACCTGTTTACCCCTATGATGTTTGTAGTTAATGATGCCT 2212
QY 1558 TCAAAAGAGCTTCAGCCGACCGTAATGACAAACAGCTCAACTATCAGTTTGTGTTAAT 1617
|||||
Db 2211 TCAAAAGAGCTTCAGCCGACCGTAATGACAAACAGCTCAACTATCAGTTTGTGTTAAT 2152
QY 1618 GCTTTTGTATTATAGCGGCTCTCTGCTTTAAATTTTCATTTATATGACCAATAGGAT 1677
|||||
Db 2151 GCTTTTGTATTATAGCGGCTCTCTGCTTTAAATTTTCATTTATATGACCAATAGGAT 2092
QY 1678 TCTTCTAGGAACCTTATCGTCTCTTTATATGACCAACCTCTGTTGTTGTTTATGTCGCG 1737
|||||
Db 2091 TCTTCTAGGAACCTTATCGTCTCTTTATATGACCAACCTCTGTTGTTGTTTATGTCGCG 2032
QY 1738 TAAAGAAATCGCTCAAAATAGTACCGTTAAACTTAATCTAACCTGTAGCAATATAAAA 1797
|||||
Db 2031 TAAAGAAATCGCTCAAAATAGTACCGTTAAACTTAATCTAACCTGTAGCAATATAAAA 1972
QY 1798 ATCTCTCTTTGGGACCTTATGTCCTCAAGGCCCTGTGTTAAATTTATGACAAATTCAG 1857
|||||
Db 1971 ATCTCTCTTTGGGACCTTATGTCCTCAAGGCCCTGTGTTAAATTTATGACAAATTCAG 1912
QY 1858 ATAATGC 1864
|||||
Db 1911 ATAATGC 1905

RESULT 2
LOCUS AP002547/2 300550 bp DNA linear BCT 25-MAY-2002
DEFINITION Chlamydomophila pneumoniae J138 genomic DNA, complete sequence,
ACCESSION AP002547 AB033786 AB033787 AB033816 AB033817 BA000008
VERSION AP002547.2 GI:10176693
KEYWORDS Chlamydomophila pneumoniae J138 (strain:J138) DNA.
SOURCE Chlamydomophila pneumoniae J138
ORGANISM Bacteria; Chlamydiales; Chlamydiaceae; Chlamydomophila.
REFERENCE 1 Shirai,M., Hirakawa,H., Ouchi,K., Tabuchi,M., Kishi,F., Kimoto,M.,
Takeuchi,A., Nishida,J., Shibata,K., Fujinaga,R., Yoneda,H.,
Matsushima,H., Tanaka,C., Furukawa,S., Miura,K., Nakazawa,A.,
Ishii,K., Shiba,T., Hattori,M., Kuhara,S. and Nakazawa,T.
Comparison of outer membrane protein genes omp and pmp in the whole
genome sequences of Chlamydia pneumoniae isolates from Japan and
the United States
J. Infect. Dis. 181 Suppl 3, S524-S527 (2000)
20298986
AUTHORS Shirai,M., Hirakawa,H., Kimoto,M., Tabuchi,M., Kishi,F., Ouchi,K.,
Shiba,T., Ishii,K., Hattori,M., Kuhara,S. and Nakazawa,T.
TITLE Comparison of whole genome sequences of Chlamydia pneumoniae J138
from Japan and CWL029 from USA
JOURNAL Nucleic Acids Res. 28 (12), 2311-2314 (2000)
MEDLINE 2030349
AUTHORS Shirai,M.
TITLE Direct Submission
JOURNAL Submitted (04-JUL-2000) Mutsunori Shirai, Yamaguchi University
School of Medicine, Department of Microbiology; 1-1-1
Minamikogushi, Ube, Yamaguchi 755-8505, Japan
(E-mail:mshirai@po.cc.yamaguchi-u.ac.jp, Tel:81-836-22-2227,
Fax:81-836-22-2415)
COMMENT On or before Sep 15, 2000 this sequence version replaced
gi:6172298, gi:6172300, gi:6172396, gi:6172398, gi:8978889.
FEATURES
Location/Qualifiers
1. 300550
/organism="Chlamydomophila pneumoniae J138"
/strain="J138"
/db_xref="taxon:138677"
/note="synonym:Chlamydia pneumoniae (strain J138)"
complement(21..794)
/gene="dapF"
source
gene
complement(21..794)
/gene="dapF"
/codon_start=1
/transl_table=11
/product="diaminopimelate epimerase"
/protein_id="BAA98725.1"
/db_xref="GI:8978890"
/translation="MAFYSPSTISKYFIYSGAGNPLIGETLPE
GFLYKPSGCAQALIIIFNSDGRPTMCGNLRCAIAHLASQKGS
GIFYSDRVLVDMTADWRAASVHRLESRPDPLPKEVVCITHTGVPHAV
SILGPEFLRYHOTFSFGVNVNFQILHCHQLRVRTYERGVEGETAACCTG
SNYSWKBSIQIHTWGGLMTVSNQRGRVYLOGSVTRDL"
complement(763..1338)
/gene="clpP_1"
complement(763..1338)
/gene="clpP_1"
/codon_start=1
/transl_table=11
/product="CLP protease"
/protein_id="BAA98726.1"
/db_xref="GI:8978891"
/translation="NADGSHKLRDIIIEKELLEARRVFSEPVTEKSADAIKMLWYL
ELKDPGKPIVFVINSPPGSDVAGFAWDOI KMLTSPVTTVVTGLAASMGVLSLCAAP
GRFATPHSRIMIHOPISIGSPITGOATDIDIHAREILTKARIIDYVVEATNQPRDI
EXAIDEDMWTANEAKDFGLDGLFSFNDL"
complement(1358..2917)
/gene="glyA"
complement(1358..2917)
/gene="glyA"
/codon_start=1
/transl_table=11
/product="serine hydroxymethyltransferase"
/protein_id="BAA98727.1"
/db_xref="GI:8978892"
/translation="MLKVPEKFKFAIVEIFTKVAVVSLHLKFLNAGSKGQSGLAS
TAYLAALHLLNAPFSGIRIGRIIDELKSQRSLKMIASENYSLSVLAMGNLLTDKYC
EGSPFKRFYSCCENDAIEWECETAKELFPAADCACVQPHSGADANLLAVMAILTHKY
QGPVSKGKGTNLTETBEETLLKAEKSCVCLGSLNSGGHLLTGNVRLNVMKLM
RCFVDVNDPTDECFOYAEISRLAKYKPKVLVAGYSSYRRLNFAVLKQIAEDCGSVL
WYDMAHFAVLGAGGVDEENPIPVADIVTTTHKTLRGPRGLVLAETREYESLTKKA
CPLMGGPLPHVIAAKTALKEALSVDFFKVAHVNNARLAEARLSHGLRLTLTGT
DNHMMVIDGLSIGISGKIAEDILSSVGIAVNRSNLSPSDAIGKWDTSIGRLGTALPTL
GNGIDMESVADIIYKVLNRLSLCHVEGSSKNGKELPEAIAQEARDRVNLLRFP
LTFEIDLEALV"
3107..3778
/gene="CPj0522"
3107..3778
/gene="CPj0522"
/codon_start=1
/transl_table=11
/product="C7433 hypothetical protein"
/protein_id="BAA98728.1"
/db_xref="GI:8978893"
/translation="MTLYIGLNQKTARKYQAHVLPILTLFPYAKSTPONKRALQELPQ
ATHVILTSSTHLSRLSMTSLSKATLTKYLCIGESTKERLLSFLGQVKVVAIQ
EIAEGIFPLQALPSSARILYPHSSLARPIREFLYNRTFFSYPHYTVKPKLKKNI
LSYKXIIIFTSPVNAFAKIPRPEKTYWCGRMTLOEFQKFSQKQVSLLETJLK
SRTSP"
3843..4175
/gene="CPj0523"
3843..4175
/gene="CPj0523"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAA98729.1"
/db_xref="GI:8978894"
/translation="MASSATPGDGTAPSLFPPTAPRPNFKALFVTTAIALVWIAL
IATTAIGLCIHPLCSFIPLTAIFYLSRYICSHVARNYIALDVVPDPHSLQDMRS
HSPIFSDR"
4223..5302
/gene="CPj0524"
4223..5302

```

THIS PAGE BLANK (USPTO)

24: em_ph:*
 25: em_pl:*
 26: em_ro:*
 27: em_sts:*
 28: em_un:*
 29: em_vi:*
 30: em_htg_hum:*
 31: em_htg_inv:*
 32: em_htg_oth:*
 33: em_htg_mus:*
 34: em_htg_pln:*
 35: em_htg_rod:*
 36: em_htg_mam:*
 37: em_htg_vrt:*
 38: em_sy:*
 39: em_hgo_hum:*
 40: em_hgo_mus:*
 41: em_hgo_oth:*

Pred. No. is the number of results predicted by chance to have a

ALIGNMENTS

RESULT 1
 AE001641/c
 LOCUS 14134 bp DNA linear BCT 01-DEC-2000
 DEFINITION Chlamydia pneumoniae section 57 of 103 of the complete genome.
 ACCESSION AE001641 AE001363
 VERSION AE001641.1 GI:4376855
 KEYWORDS
 SOURCE Chlamydophila pneumoniae CWL029.
 ORGANISM Chlamydophila pneumoniae CWL029
 Bacteria; Chlamydiales; Chlamydiaceae; Chlamydophila.
 REFERENCE 1 (bases 1 to 14134)
 AUTHORS Kalman,S., Mitchell,W., Marathe,R., Lamel,C., Fan,J., Hymn,R.W.,
 Olinger,L., Grimwood,J., Davis,R.W. and Stephens,R.S.
 TITLE Comparative genomes of Chlamydia pneumoniae and C. trachomatis
 JOURNAL Nat. Genet. 21 (4), 385-389 (1999)

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 17, 2002, 06:18:31 ; Search time 3364 Seconds
(without alignments)
16125.930 Million cell updates/sec

Title: US-09-868-987-1
Perfect score: 1864
Sequence: 1 atggacttcgcattgtc.....atgacaaattcagataatgc 1864

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

- 1: gb_ba.*
- 2: gb_htg.*
- 3: gb_in.*
- 4: gb_on.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vi.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_or.*
- 21: em_ov.*
- 22: em_pat.*
- 23: em_ph.*
- 24: em_pi.*
- 25: em_ro.*
- 26: em_sts.*
- 27: em_un.*
- 28: em_vi.*
- 29: em_vt.*
- 30: em_htg_hum.*
- 31: em_htg_inv.*
- 32: em_htg_other.*
- 33: em_htg_mus.*
- 34: em_htg_pln.*
- 35: em_htg_rod.*
- 36: em_htg_mam.*
- 37: em_htg_vrt.*
- 38: em_sy.*
- 39: em_btgo_hum.*
- 40: em_btgo_mus.*
- 41: em_btgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
C 1	1829.4	98.1	14134	1	AE001641
C 2	1829.4	98.1	300550	1	AE002547
C 3	1797.4	96.4	12198	1	AE002179
C 4	773.6	41.5	9989	1	AE001318
C 5	760.8	40.8	12442	1	AE002341
C 6	161.2	8.6	11527	1	AE010580
C 7	153.6	8.2	119836	1	AE001567
C 8	148.2	8.0	282183	1	CJ11168X4
C 9	146.4	7.9	15058	1	AE000652
C 10	144.8	7.8	301550	1	AE003192
C 11	127	6.8	12592	1	AE007729
C 12	124	6.7	300600	1	AP005369
C 13	122.4	6.6	1557	1	AY078352
C 14	122	6.5	12504	1	AE001219
C 15	119.6	6.4	279110	1	RPX03
C 16	119.4	6.4	301550	1	AP003134
C 17	119.4	6.4	333750	1	AP004827
C 18	119.4	6.4	342600	1	AP003363
C 19	118	6.3	45613	6	AX067447
C 20	115	6.2	10963	1	AE013703
C 21	115	6.2	205050	1	AJ414155
C 22	112.6	6.0	10899	1	AE008643
C 23	110.2	5.9	13990	1	AE001166
C 24	109.6	5.9	11540	1	AE009509
C 25	108.8	5.8	315079	1	MPULM03
C 26	108.2	5.8	1848	6	AX084084
C 27	108.2	5.8	1848	6	AX395145
C 28	108.2	5.8	13818	1	U32710
C 29	106.4	5.7	349619	1	AP002996
C 30	103.6	5.6	349980	6	AX413017
C 31	103.6	5.6	349980	6	AX417045
C 32	102.8	5.5	12201	1	AE006057
C 33	102.2	5.5	3000	1	RCU89979
C 34	102.2	5.5	5839	6	AX416793
C 35	102	5.5	240050	1	AL591979
C 36	100.6	5.4	7687	1	AE012971
C 37	100	5.4	14261	1	AE004160
C 38	100	5.4	14567	1	AE000716
C 39	99.6	5.3	2226	6	AX142455
C 40	99.6	5.3	3511	1	AF269419
C 41	99.6	5.3	3511	6	AX144739
C 42	99.6	5.3	3705	1	AF269550
C 43	99.6	5.3	3705	6	AX144870
C 44	99	5.3	4357	1	AE013081
C 45	98.2	5.3	11217	1	AE003876

ALIGNMENTS

RESULT 1
AE001641/c
LOCUS
DEFINITION Chlamydia pneumoniae section 57 of 103 of the complete genome.
ACCESSION AE001641 AE001363
VERSION AE001641.1 GI:4376855
KEYWORDS
SOURCE Chlamydia pneumoniae CWL029
ORGANISM Chlamydia pneumoniae; Chlamydiaceae; Chlamydia.
REFERENCE 1 (bases 1 to 14134)
AUTHORS Kalman,S., Mitchell,W., Marathe,R., Lammel,C., Fan,J., Hyman,R.W.,
Olinger,L., Grimwood,J., Davis,R.W. and Stephens,R.S.
TITLE Comparative genomes of Chlamydia pneumoniae and C. trachomatis
JOURNAL Nat. Genet. 21 (4), 385-389 (1999)

MEDLINE
 99206606
PUBMED
 2 (bases 1 to 14134)
REFERENCE
 Kaiman, S., Mitchell, W., Marathe, R., Lammel, C., Fan, J., Olinger, L.,
 Grimwood, J., Davis, R. W. and Stephens, R. S.
AUTHORS
 Submitted (01-DEC-1998) Program in Infectious Diseases, University
 of California, 235 Earl Warren Hall, Berkeley, CA 94720, USA
TITLE
 Direct Submission
JOURNAL
 Location/Qualifiers
FEATURES
 source
 1. .14134
 /organism="Chlamydia pneumoniae CWL029"
 /strain="CWL029"
 /db_xref="taxon:115713"
 /complement (158. .1924)
 /gene="recJ"
 /note="Cp0563"
 /complement (158. .1924)
 /gene="recJ"
 /codon_start=1
 /transl_table=11
 /product="ssDNA Exonuclease"
 /protein_id="AAD18703.1"
 /db_xref="GI:4376856"
 /translation="MTNSDNASAAAGLMAHKEDPAFLGMIKEPHLPPTVAQIFSR
 GPTQIPIHKFLYSHSLYDPCLEFDMGKAVSRILARPKGHMIVGDDQDMG
 VALLVEFLRIDIVHVSFYFLGALILKQGETSTILAKKEGITLITVDCGTRAGEV
 SDITRQIGIDVITDHNMPGKIIPHCATINPKLRDHYPERELTGVGVARGLAGVINK
 ALISRNLPVPSQSLKKLLDLVLTGITDGVLLGENRVVVRGIGKEIARFAGRLNKL
 LCAICGVESEVTSTDIVLKIAPKLNLSGRLDPAKGVELLITQDERVALIMEDN
 INNRORIEAEVODVOEILNSNPETLKOAYILASPMARAVIPISALATYKND
 VVITAIORIGKGSARTIGSPFLIGTKKSSILISGCHDPAAGVIMKEDKEDKK
 KPVHLVNSLSKGGDTLPHEIDAIADDAIDYDLASMEIFEPFGKINLMPITYSKYR
 QVAPYKVLPGNHLKLYLSQKERNLEGVAFGLGRHADLKASWHYPLEIATYPLSOTS
 GSGCVHLVDFRIFISEPRFSD"
 /complement (2010. .6218)
 /gene="secD/secF"
 /note="Cp0564"
 /complement (2010. .6218)
 /gene="secD/secF"
 /codon_start=1
 /transl_table=11
 /product="Protein Export Proteins SecD/SecF (fusion)"
 /protein_id="AAD18704.1"
 /db_xref="GI:4376857"
 /translation="MKQVKRNFALIIICVPAALAYVLPCLYYAKPLDKKIDGENE
 HIILKFTKQAOVKDVIPIVSAIILSLHRLGHIOQHPAIDIVSARFKGBDAEPTI
 GNLVHGEPNPIKSARLHVGVYSGREHDHIVQVSAINTSLVESDFSVASENEED
 MASSIIQRYVSACTPPKQDCSCSYPIETWETAREOLQAKNLSGSEFVSRLSAP
 COQSPSSNQRLAFLRSLSSLDMAIDVDQKLKRYETLSQTAIRSLDCPYIEG
 LRLDCSESSLPFSISIEYCPKERRKIPLTILHSDLAQRTSLSKQRLDPSRLAYEQKL
 SKULTVQVEDYNNGSFQWMDKDTQKIILOGERLLQGLAEHLTALTLHRAESCDL
 IPIENFVFCRQPRESEAFQCYIFSPNDCGSGSVYIILKGLRSIVAKYQOGGGE
 LOSFEKDLQNLVNCFSHTEAISWTLGEOVLEIRHPLQOGLDVWGEVFLGKGCGL
 EVDKIDQRLATVNOIEKNROSDLVRMHEOYRAKCSMDLERISAPIPYONLPLEMK
 LMKRPSRGENTIRLIGIDPYGROQLISFKDHQKOLITJEDKILKYSDEICARLNLK
 VSEIEKRESDYIHLISVSGSTISSSTILSTSKSRFVNERSSYASRYEYQRLD
 YLWFTSOAQKTSPEIRINTASALFNEEVVPPSVHEATIKLSSEGLAFSPSCETPS
 TDLDTPTSMIAIGDAOKANPLVIYVRNALYASLIDIRPFPAAGEGVILNFSPL
 TSPKRAEKLSPTESFHTWTSAVCOEGISGTANGQYANRGMMAVVIDGMVSSPL
 NVPLKHNASVSGKTRHEVGLASDLKSGMSRPEVLESETTSSDLGKOCOCIGIIS
 ACCGIALMLIYMSYTRFGVYASGAVLMLLILWALQYLDAPLTLSGLAGVILMG
 MAVDANLVVERIREEFLLSQSLKSEYKGYTAFGAIIDSNLITVIALSLFFLDTG
 PIKGFALTLLIGIFSSMFTALFMTKFFPMLMMNKTOHTOLAMNKFVGIGDLRGCK
 PLMAVSGSVFLGLGVALGFGAMNSVLMGDFRGYAFENPKHEGIDVAQMRKVAHK
 LOEGASSRDFRIQTFGSSERKIYFSDKALSTKADTSLPKINDEHLAVALSE
 TGLDSEPTENQNMFSKVSLSKSMRYOATIGLIGALAIILLYSVLAFEMOYAS
 AVCLLHDLATCAVLEFIARFLKTIQIDIGALMTVYGVISANTLITFDPIRER
 QANLFTPMHVLVNDALOKTFSRTVMTATTLVLMMLFTGSSSVFNFATIMTIGILL
 GTLSLIYAPPLLFLFVRKENSRLK"
 /gene="Cp0565"
 /complement (6398. .7498)
 /gene="Cp0565"
 /complement (6398. .7498)
 /gene="Cp0565"
 /complement (6398. .7498)

/codon_start=1
 /transl_table=11
 /product="Ct49 hypothetical protein"
 /protein_id="AAD18705.1"
 /db_xref="GI:4376858"
 /translation="MKPPIISIDLIPTQVIMWRBGGNVHYVPAONLPEKILGVL
 ACRGALIGCAAFAGVCOITIPCIIGMILIGILGPAILOYSKMSRERLRFETK
 VEEKPIINMGCSLISQSMKIRPGCYIHGCGYQVYICSSQSEIVTKIRPKSDRNTSI
 FLIQEDQIALKQGIKESLSKRTAIDPSVSSVLSLEIQREEQQLDPKVISMSSED
 QASDRTHPKSAIYVNI SDAAQEPQRCYIDATYKAFVFLDQIDGPNIVKHTIYLT
 PLIGVDALPKKEQENMLKLSQAAFLVSAEYAKRKEKODSIRIKIFITDPTSPTS
 LYFSPHSSTPSVTPISLGSFVGESESTFA"
 8009. .8761
 /gene="yaeS"
 /note="Cp0566"
 8009. .8761
 /gene="yaeS"
 /codon_start=1
 /transl_table=11
 /product="yaeS family hypothetical protein"
 /protein_id="AAD18706.1"
 /db_xref="GI:4376859"
 /translation="MSLATNNASKEPSPQLRPNHVAIIMDGNRRMYKKRRECGHTH
 TSGHYAGKAVLNINAAVLDLGIKVLITVTFSTENGRKKEIOELFNIPTYOLDKOL
 PLYMENEICLRIGDLSKLPKGIOTKINNVSMRTASFSLVILVAVNGKDEILRAF
 KKLHVDILNKRTISSDLSLSLYLDLSTGLDPPDLLITFGSEKMSVSNFLMQIAYTE
 LITITDLMFDFPQDLFEATINYIQRRRGK"
 8762. .9688
 /gene="cdaA"
 /note="Cp0567"
 8762. .9688
 /gene="cdaA"
 /codon_start=1
 /transl_table=11
 /product="Phosphatidate Cytidylyltransferase"
 /protein_id="AAD18707.1"
 /db_xref="GI:4376860"
 /translation="MLNSKRFKSGKAGVDLQRYVNVHSLVTPVLVLLYSLEPLTS
 FALGFTLTCGAVGTREYSKMAKAKMHPPLSTFSLGSLFLASLIRKMSLDFG
 FDLFWTLTIWVWVSIFRVRKSTIGALDSVTLSEIIVYVPIRLFLVLSYFHT
 QEPYLGIMWASFLIATTKGADIFGVFFGAFQNKXIAPDISNKTIVGFVACCLGATL
 ISGIFFLQPTREFASFPWPAIILPLGLALGITGFFGDIIESIFKRDALKNKNKLA
 VGMDLDTLSLSTPIATVFLITQSKFIC"
 9685. .10335
 /gene="cmk"
 /note="Cp0568"
 9685. .10335
 /gene="cmk"
 /codon_start=1
 /transl_table=11
 /product="CMP Kinase"
 /protein_id="AAD18708.1"
 /db_xref="GI:4376861"
 /translation="MIRITDGSSTGKSTTAKALADHLHPNYCNGKMYRTALVARLO
 SPWATLPLTKLEBPESFTFATGQPLSEFNGHLTSETLTOEVANASSELSDPEV
 RAKPMQDLQRRVQALNCVPEGRDMGSKVPPNADIKITFLTSSPEVAPQARKULDPBETL
 SEBQQAELVKKRQADACRAHPPLVIPENGIVDSDDLITROYLEKITALFNNEL"
 10332. .10970
 /gene="pisc"
 /note="Cp0569"
 10332. .10970
 /gene="pisc"
 /codon_start=1
 /transl_table=11
 /product="Glycerol-3-P Acyltransferase"
 /protein_id="AAD18709.1"
 /db_xref="GI:4376862"
 /translation="MIRICKFTWVWASLFLKLVYGVYKYNKPIGAPALIAVNNHSL
 DPLAHMCVHCETIHLRASLFINIPLMWQWOCFPRQDEGSAFKAISRLFNKSKK
 LVIYEPGASPDGLOPKGVGIGMAAKSRVPIIPVYIRGTAEAFNRHQIKIPVWXTI
 TCVFGTPMYFDIIONPEIKNKETVQIITNQTMNKIABLKAWEYSGCKGDVP"
 10975. .12660
 /gene="args"

CDS	gene	BASE COUNT	4180 a	2981 c	2805 g	4168 t	ORIGIN
							/note="CPn0570" 10975..12660 /gene="argS" /codon_start=1 /transl_table=11 /product="Arginyl tRNA Transferase" /protein_id="AADI8710.1" /db_xref="GI:4376863" /translation="MSTLSILSVISQAIKAPPNLEDAWPEITTSKEHFHYOCN DAMKLARVLKAPRAIAEAIVASLPQSPFSLIAGAGFINTFSPVFLNQOLEHFKD ALKLGQVQSPKKIIIDFSSPNIAKDMHVGLHRLSTIIIGSLARIISVVGHDVLRNHI GDMGTAFGLMITYLQENPCDYSLEDLTSLYKKAYCFYNDEEFKRSQONVVALQAK DPQAIWIKICTSKAFQKIYDIIIVVEKRGESFYNPFLEIILEDLEKGLLTVS NDAKCVFHEAFSPFWQSGDGYVATTDLAAMRYIIRIBEDHAKKIIVTDLGOSLHF QLEEAITAAGLYQPLFSGHGVGLVDPOGKKLKRSGENVKLRELLDTAIEKAEFA RHERELDTAQERAPIYINAIKYSDLSHRTSDYVFSPEKMLFEENTAMFLLY AVVRIGIQIKRRLGISOLSEPGPPEIOEPAEELIALTLIRPEALESTIKELCPHFLT YLYNLTHKNGFRDRSHIQDSPYAKSRSLFCALAEQVLATGMHLLGLKTLRL" complement(12614..13990) /gene="murA" /note="CPn0571" complement(12614..13990) /gene="murA" /codon_start=1 /transl_table=11 /product="UDP-N-Acetylglucosamine Transferase" /protein_id="AADI8711.1" /db_xref="GI:4376864" /translation="MOIAQVFCGRNLNGEVKVSAGKNAATKLIVASLLSDOKTCLRV PDIGVSLTVELCKSLGAHVSWDKETEVLIIYPTQCTRVPTFSNVNPIPLLIGA HLKCPGEGVPTVGGDAIGERTLNFHFEGLKQGVQISDSGSGYAKAPRGUKGYDI LHPYVSAGTENIIAAIHAKGRTVIKNVALEAEILDLVLFQAGADITDNDRTYI FIVTGIGLSVDHPIPLDKTEAASFGMAVVGSGRVFVNAKQELLIPFLKMLRSIGG FLVSGSIFBFQSRPLVGVGLVTDVHPGLTDQOQPFVALLSQAGSSVIHETVHEN RLGYLHLOHMGAEACOLFHQCLSKACRYAIGNPFSAVIHGATPLWASHLVPDLRA GFAYNWALIAEGGGSIENHLLDRGYINWVKLSLGAKIQIFDMEQEELITSPKS LAURDASL" complement(12614..13990) /gene="murA" /note="CPn0571" complement(12614..13990) /gene="murA" /codon_start=1 /transl_table=11 /product="UDP-N-Acetylglucosamine Transferase" /protein_id="AADI8711.1" /db_xref="GI:4376864" /translation="MOIAQVFCGRNLNGEVKVSAGKNAATKLIVASLLSDOKTCLRV PDIGVSLTVELCKSLGAHVSWDKETEVLIIYPTQCTRVPTFSNVNPIPLLIGA HLKCPGEGVPTVGGDAIGERTLNFHFEGLKQGVQISDSGSGYAKAPRGUKGYDI LHPYVSAGTENIIAAIHAKGRTVIKNVALEAEILDLVLFQAGADITDNDRTYI FIVTGIGLSVDHPIPLDKTEAASFGMAVVGSGRVFVNAKQELLIPFLKMLRSIGG FLVSGSIFBFQSRPLVGVGLVTDVHPGLTDQOQPFVALLSQAGSSVIHETVHEN RLGYLHLOHMGAEACOLFHQCLSKACRYAIGNPFSAVIHGATPLWASHLVPDLRA GFAYNWALIAEGGGSIENHLLDRGYINWVKLSLGAKIQIFDMEQEELITSPKS LAURDASL" complement(12614..13990) /gene="murA" /note="CPn0571" complement(12614..13990) /gene="murA" /codon_start=1 /transl_table=11 /product="UDP-N-Acetylglucosamine Transferase" /protein_id="AADI8711.1" /db_xref="GI:4376864" /translation="MOIAQVFCGRNLNGEVKVSAGKNAATKLIVASLLSDOKTCLRV PDIGVSLTVELCKSLGAHVSWDKETEVLIIYPTQCTRVPTFSNVNPIPLLIGA HLKCPGEGVPTVGGDAIGERTLNFHFEGLKQGVQISDSGSGYAKAPRGUKGYDI LHPYVSAGTENIIAAIHAKGRTVIKNVALEAEILDLVLFQAGADITDNDRTYI FIVTGIGLSVDHPIPLDKTEAASFGMAVVGSGRVFVNAKQELLIPFLKMLRSIGG FLVSGSIFBFQSRPLVGVGLVTDVHPGLTDQOQPFVALLSQAGSSVIHETVHEN RLGYLHLOHMGAEACOLFHQCLSKACRYAIGNPFSAVIHGATPLWASHLVPDLRA GFAYNWALIAEGGGSIENHLLDRGYINWVKLSLGAKIQIFDMEQEELITSPKS LAURDASL" complement(12614..13990) /gene="murA" /note="CPn0571" complement(12614..13990) /gene="murA" /codon_start=1 /transl_table=11 /product="UDP-N-Acetylglucosamine Transferase" /protein_id="AADI8711.1" /db_xref="GI:4376864" /translation="MOIAQVFCGRNLNGEVKVSAGKNAATKLIVASLLSDOKTCLRV PDIGVSLTVELCKSLGAHVSWDKETEVLIIYPTQCTRVPTFSNVNPIPLLIGA HLKCPGEGVPTVGGDAIGERTLNFHFEGLKQGVQISDSGSGYAKAPRGUKGYDI LHPYVSAGTENIIAAIHAKGRTVIKNVALEAEILDLVLFQAGADITDNDRTYI FIVTGIGLSVDHPIPLDKTEAASFGMAVVGSGRVFVNAKQELLIPFLKMLRSIGG FLVSGSIFBFQSRPLVGVGLVTDVHPGLTDQOQPFVALLSQAGSSVIHETVHEN RLGYLHLOHMGAEACOLFHQCLSKACRYAIGNPFSAVIHGATPLWASHLVPDLRA GFAYNWALIAEGGGSIENHLLDRGYINWVKLSLGAKIQIFDMEQEELITSPKS LAURDASL" complement(12614..13990) /gene="murA" /note="CPn0571" complement(12614..13990) /gene="murA" /codon_start=1 /transl_table=11 /product="UDP-N-Acetylglucosamine Transferase" /protein_id="AADI8711.1" /db_xref="GI:4376864" /translation="MOIAQVFCGRNLNGEVKVSAGKNAATKLIVASLLSDOKTCLRV PDIGVSLTVELCKSLGAHVSWDKETEVLIIYPTQCTRVPTFSNVNPIPLLIGA HLKCPGEGVPTVGGDAIGERTLNFHFEGLKQGVQISDSGSGYAKAPRGUKGYDI LHPYVSAGTENIIAAIHAKGRTVIKNVALEAEILDLVLFQAGADITDNDRTYI FIVTGIGLSVDHPIPLDKTEAASFGMAVVGSGRVFVNAKQELLIPFLKMLRSIGG FLVSGSIFBFQSRPLVGVGLVTDVHPGLTDQOQPFVALLSQAGSSVIHETVHEN RLGYLHLOHMGAEACOLFHQCLSKACRYAIGNPFSAVIHGATPLWASHLVPDLRA GFAYNWALIAEGGGSIENHLLDRGYINWVKLSLGAKIQIFDMEQEELITSPKS LAURDASL" complement(12614..13990) /gene="murA" /note="CPn0571" complement(12614..13990) /gene="murA" /codon_start=1 /transl_table=11 /product="UDP-N-Acetylglucosamine Transferase" /protein_id="AADI8711.1" /db_xref="GI:4376864" /translation="MOIAQVFCGRNLNGEVKVSAGKNAATKLIVASLLSDOKTCLRV PDIGVSLTVELCKSLGAHVSWDKETEVLIIYPTQCTRVPTFSNVNPIPLLIGA HLKCPGEGVPTVGGDAIGERTLNFHFEGLKQGVQISDSGSGYAKAPRGUKGYDI LHPYVSAGTENIIAAIHAKGRTVIKNVALEAEILDLVLFQAGADITDNDRTYI FIVTGIGLSVDHPIPLDKTEAASFGMAVVGSGRVFVNAKQELLIPFLKMLRSIGG FLVSGSIFBFQSRPLVGVGLVTDVHPGLTDQOQPFVALLSQAGSSVIHETVHEN RLGYLHLOHMGAEACOLFHQCLSKACRYAIGNPFSAVIHGATPLWASHLVPDLRA GFAYNWALIAEGGGSIENHLLDRGYINWVKLSLGAKIQIFDMEQEELITSPKS LAURDASL" complement(12614..13990) /gene="murA" /note="CPn0571" complement(12614..13990) /gene="murA" /codon_start=1 /transl_table=11 /product="UDP-N-Acetylglucosamine Transferase" /protein_id="AADI8711.1" /db_xref="GI:4376864" /translation="MOIAQVFCGRNLNGEVKVSAGKNAATKLIVASLLSDOKTCLRV PDIGVSLTVELCKSLGAHVSWDKETEVLIIYPTQCTRVPTFSNVNPIPLLIGA HLKCPGEGVPTVGGDAIGERTLNFHFEGLKQGVQISDSGSGYAKAPRGUKGYDI LHPYVSAGTENIIAAIHAKGRTVIKNVALEAEILDLVLFQAGADITDNDRTYI FIVTGIGLSVDHPIPLDKTEAASFGMAVVGSGRVFVNAKQELLIPFLKMLRSIGG FLVSGSIFBFQSRPLVGVGLVTDVHPGLTDQOQPFVALLSQAGSSVIHETVHEN RLGYLHLOHMGAEACOLFHQCLSKACRYAIGNPFSAVIHGATPLWASHLVPDLRA GFAYNWALIAEGGGSIENHLLDRGYINWVKLSLGAKIQIFDMEQEELITSPKS LAURDASL" complement(12614..13990) /gene="murA" /note="CPn0571" complement(12614..13990) /gene="murA" /codon_start=1 /transl_table=11 /product="UDP-N-Acetylglucosamine Transferase" /protein_id="AADI8711.1" /db_xref="GI:4376864" /translation="MOIAQVFCGRNLNGEVKVSAGKNAATKLIVASLLSDOKTCLRV PDIGVSLTVELCKSLGAHVSWDKETEVLIIYPTQCTRVPTFSNVNPIPLLIGA HLKCPGEGVPTVGGDAIGERTLNFHFEGLKQGVQISDSGSGYAKAPRGUKGYDI LHPYVSAGTENIIAAIHAKGRTVIKNVALEAEILDLVLFQAGADITDNDRTYI FIVTGIGLSVDHPIPLDKTEAASFGMAVVGSGRVFVNAKQELLIPFLKMLRSIGG FLVSGSIFBFQSRPLVGVGLVTDVHPGLTDQOQPFVALLSQAGSSVIHETVHEN RLGYLHLOHMGAEACOLFHQCLSKACRYAIGNPFSAVIHGATPLWASHLVPDLRA GFAYNWALIAEGGGSIENHLLDRGYINWVKLSLGAKIQIFDMEQEELITSPKS LAURDASL" complement(12614..13990) /gene="murA" /note="CPn0571" complement(12614..13990) /gene="murA" /codon_start=1 /transl_table=11

Db	2271	TCGTGAAGATGCCAGGAACCTGTTTACCCCTATGCATGCTTTAGTATGANGCCCT	2212	CDS	complement (21..794) /gene="dapF"
Qy	1558	TCAAAAGACGTTCAAGCCGACGATATGACAAACGCTACACTATCACTTTGTTAAT	1617		/codon_start=1 /transl_table=11 /product="diaminopimelate epimerase"
Db	2211	TCAAAAGACGTTCAAGCCGACGATATGACAAACGCTACACTATCACTTTGTTAAT	2152		/protein_id="BAA98725.1"
Qy	1618	GCTTTTGGTTTATAGCGCGCTCTCGTCTTTAATTTTGCATTTTATAGACATAGGGAT	1677		/db_xref="GI:8978890"
Db	2151	GCTTTTGGTTTATAGCGCGCTCTCTCGTCTTTAATTTTGCATTTTATAGACATAGGGAT	2052		/translation="MAFVPSPTISKVFYISGAGNRFLIGETLPEVEDVPRFLGQETRD GFLYLKPPSCADAQILIRSDSSRPTMGNGRCALAHLASQKGSIDSISVSDSGIS GIFYSDRVLVDMTADWPCASVHRLSRPDLPKKVCVCHTGVPAHVLVLPILSTLD SLIGPELRHVOTFSDGVNVNFVQILGHCOLARRYERGVBEETAACGALASALV SNSYKESIQHTWGEMLTVSQNRGRVYLQGSVTRDL"
Qy	1678	TCTTCTAGGAAGCTTATAGTCTCTTTATATATGACACACTCTGTGTGTTTATAGTCG	1737	gene	complement (763..1338) /gene="c1p1"
Db	2091	TCTTCTAGGAAGCTTATAGTCTCTTTATATATGACACACTCTGTGTGTTTATAGTCG	2032	CDS	complement (763..1338) /gene="c1p1"
Qy	1738	TAAAGAAATCGCTCAAAATAGTACCGTTAACTTAATCTAGCGTGAATATATAA	1797		/codon_start=1 /transl_table=11 /product="C1P1 protease"
Db	2031	TAAAGAAATCGCTCAAAATAGTACCGTTAACTTAATCTAGCGTGAATATATAA	1972		/protein_id="BAA98726.1"
Qy	1798	ATCTCTTGGGACTTATAGTCCCAAGGCCCTGTGTATTAATTAATGACAAATTCAG	1857		/db_xref="GI:8978891"
Db	1971	ATCTCTTGGGACTTATAGTCCCAAGGCCCTGTGTATTAATTAATGACAAATTCAG	1912		/translation="MADGVHKLRLDIEKELLBARVFESEPTKSDAISKUWYL ELKDPKPIVFNINSFGSVDAFVMDQIKMLTSPVTIVVGLAASMGSVSLCAP GRFATPHSRIMIHQPSIGSPITGQATDLDIHAREILTKTARIIDVYVATNQPDII EXAIDRDMMTANBEAKDFGLDGLIFSFNDL"
Qy	1858	ATAATGC 1864		gene	complement (1358..2917) /gene="glpA"
Db	1911	ATAATGC 1905		CDS	complement (1358..2917) /gene="glpA"
RESULT 2	AP002547	300550 bp	DNA		/codon_start=1 /transl_table=11 /product="serine hydroxymethyltransferase"
LOCUS	AP002547/c				/protein_id="BAA98727.1"
DEFINITION	Chlamydomophila pneumoniae J138 genomic DNA, complete sequence,				/db_xref="GI:8978892"
ACCESSION	AB002547	AB033786	AB033787	AB033816	BA000008
VERSION	AP002547.2	GI:10176693			
KEYWORDS	Chlamydomophila pneumoniae J138 (strain:J138) DNA.				
SOURCE	Bacteria; Chlamydiales; Chlamydiaceae; Chlamydomophila.				
ORGANISM	Chlamydomophila pneumoniae J138				
REFERENCE	1				
AUTHORS	Shirai, M., Hirakawa, H., Ouchi, K., Tabuchi, M., Kishi, F., Kimoto, M., Takeuchi, A., Nishida, J., Shibata, K., Fujinaga, R., Yoneda, H., Matsushima, H., Tanaka, C., Furukawa, S., Miura, K., Nakazawa, A., Ishii, K., Shiba, T., Hattori, M., Kubera, S. and Nakazawa, T.				
TITLE	Comparison of outer membrane protein genes omp and pmp in the whole genome sequences of Chlamydia pneumoniae isolates from Japan and the United States				
JOURNAL	J. Infect. Dis. 181 Suppl 3, S524-S527 (2000)				
MEDLINE	20298986				
REFERENCE	2				
AUTHORS	Shirai, M., Hirakawa, H., Kimoto, M., Tabuchi, M., Kishi, F., Ouchi, K., Shiba, T., Ishii, K., Hattori, M., Kuhara, S. and Nakazawa, T.				
TITLE	Comparison of whole genome sequences of Chlamydia pneumoniae J138 from Japan and CML029 from USA				
JOURNAL	Nucleic Acids Res. 28 (12), 2311-2314 (2000)				
MEDLINE	20330349				
REFERENCE	3 (bases 1 to 300550)				
AUTHORS	Shirai, M.				
JOURNAL	Direct Submission				
TITLE	Submitted (04-JUL-2000) Mutsunori Shirai, Yamaguchi University School of Medicine, Department of Microbiology; 1-1-1 Minamikogushi, Ube, Yamaguchi 755-8505, Japan (E-mail: mshirai@po.cc.yamaguchi-u.ac.jp, Tel: 81-836-22-2227, Fax: 81-836-22-2415)				
COMMENT	On or before Sep 15, 2000 this sequence version replaced gi:6172298, gi:6172300, gi:6172396, gi:6172398, gi:6978889. location/Qualifiers				
FEATURES	1..300550				
source	/organism="Chlamydomophila pneumoniae J138"				
gene	/db_xref="taxon:138677"				
	/note="synonym: Chlamydia pneumoniae (strain J138)"				
	complement (21..794)				
	/gene="dapF"				

```
/gene="CPj0524"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAA98730.1"
/db_xref="GI:8978895"
/transl_table="MSGPSSTESQSVLSVPRDKKAIKQFTIAKISTAILASL
ALGAVAGISLITVILGPNFVALIITLITLALSVVTLVYHQMTSKVSNWQKLEQNFK
PLGKAWQKNCVSNMFPYNNHLPKFAVQTDASQFPQFTFLGLRVIEKNQST
GIIFNPVGNPTNLSTILSTLYSLKDKSVMDTKCKREGGPKAGEDPFEKTEVRV
VKLPNEALDQTNLNLSSAKSKILPTFLGHVCPKSEELPNOQEVYRQALLAVENCL
KAATESHAALVALPLFTSVVEVPEEILPKEGIFYWQDQQAFCRALLDAIQNTALR
YQPSRLVILQDPNTIESQSRSE"
5642..6406
/gene="CPj0525"
5642..6406
/gene="CPj0525"
/codon_start=1
/transl_table=11
/product="CTj398 hypothetical protein"
/protein_id="BAA98731.1"
/db_xref="GI:8978896"
/transl_table="MHDALLSILAIQELDIKMLRVKKEHOKELAKVOSLKSIDRR
KVOEKELMENLKTQIEDGNRIQEISEQINKLENQAAVKMDENALIQEMTANK
ERSLEHQLSDLMQKAGGDLIVSLKESLASTENSSVIEKEIFESIKKINEGRAL
LEQTEUKHATNPELISYIERLLNNKDRVVVPIENRVSCGCHIVLTPQHENLVRRKD
RLIFCEHRSRLYQWESQVNAQENSTARRRRRRAAV"
complement(6827..7816)
/gene="kpsF"
complement(6827..7816)
/gene="kpsF"
/codon_start=1
/transl_table=11
/product="GutQ/KpsF family sugar-P isomerase"
/protein_id="BAA98732.1"
/db_xref="GI:8978897"
/transl_table="MPSPMISTDVCQDILQKQEAUVFFQAPQPKAMQLAEKILGH
SGWFFGKGVKQVARKVATLQSLSERALFFSPVDLHGLDGLSPGDIIVCLFSKS
GETQELDTVTHPKSREAIIVAITSMPSYLNLAALSDIIVLPSVAELDPNLIPTNST
TCQMI FGDILAMLLFHRGQVSLSTYKGNHPSGQVGMKANGKVDMPFKTEVPFCHLG
DKVFSLEVSAYCGCVCIVDQFRLMGITFDGDLRSLSAYSGVEFSLISLEKWTIA
NPRCITSDSIATALQWESSSPVAVLPVLVDNEENRHVTGLLHMHTLAKAGLL"
complement(7843..9024)
/gene="sucB 2"
complement(7843..9024)
/gene="sucB 2"
/codon_start=1
/transl_table=11
/product="diacylglycerol lipid succinyltransferase"
/protein_id="BAA98733.1"
/db_xref="GI:8978898"
/transl_table="MIEFPPKIGETSSGSIVRWLKNLGDHVARDEPLIEVSTDKI
ATELPSKAGRLVFCVNEGDEVASGDVLGLIELEISEADDESTSCPLTSCETKSA
GSSSSVWFSPAVLSLAQRGIGLDNLQKTAGTKGGRVTRQDLEATISQSVSIFE
IFQGVNRIEMSPURRAIASLSKSDSEVPASLVVDVDTDLNMLISGERQRFPLDH
GVLKTTISFTVQLAQRQFPLNGSLDGTIIVMKSVNVGVAVNLNKEGVVPPVTH
NCQDRGLVSTAKALADLSRARLNKLDPSVQDGSVTNFGMTGALIGMPIIRYPEV
AIIIGIGTIQKRVVRRDDSLAIRKMVVVTLTFDRHVLVDGIYGSFSLTSLKNRLESVTM
G"
Complement(9038..10282)
/gene="glt"
complement(9038..10282)
/gene="glt"
/codon_start=1
/transl_table=11
/product="glucamate symport"
/protein_id="BAA98734.1"
/db_xref="GI:8978899"
/transl_table="MKLWMKIFIGLVGVTLGLVLEDAIKFFKPIGDIIFNLISMVVY
PLVFCSMVLGIASIDMKKLGRIKIGKSVGLYLTALAIIVIGLCFAMI FSPGNGDFA
QAQMSDAVTVDSNKTAAVPLSIIAQVPPSNVRSFABGNILQIIIFAIPLGIALRL
SGEGRVPERFIDGFSIMLRVMNIMSFAPYGVGASMAWISGNHGLGVQLWQKFLI
AYYLACLHFATLVFGGLVRFECCKMSFSKFLSSMDAISCAVSTASSATLPLVTMRCVS
```

```
KNLGVSAEVSQFVLPLGATVNMNGTAIFQGMAAVFIAQAYNCPLSLSSLLLVVTATP
SAVSGAVFGGMITIGSVLASVGLPIQGIILAGIDRLRDIIVGTPMNILGDVAVTAY
VASGEGELSPYESIKQESVET"
complement(10282..11379)
/gene="ycaH"
complement(10282..11379)
/gene="ycaH"
/codon_start=1
/transl_table=11
/product="AtPase"

Query Match 98.1%; Score 1829.4; DB 1; Length 300550;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1863; Conservative 0; Mismatches 1; Indels 3; Gaps 3;

QY 1 ATGGACTTCCGCATATTGTTCAGGAGGGGATCAGCGGCACCTGCTAATGGACAATATTTCTGC 60
DB 51024 ATGGACTTCCGCATATTGTTCAGGAGGGGATCAGCGGCACCTGCTAATGGACAATATTTCTGC 50965

QY 61 AAACCGTGGATGGCGTATGGCTGTAGTGTGACGGTTATATGTCTCAGCAGCCCTATTTT 120
DB 50964 AAACCGTGGATGGCGTATGGCTGTAGTGTGACGGTTATATGTCTCAGCAGCCCTATTTT 50905

QY 121 AAACGTCCTCCATTGAAAATCATGTCAGTGTCTCAGGAAAATTTACCCACCCTGAAGTGAG 180
DB 50904 AAACGTCCTCCATTGAAAATCATGTCAGTGTCTCAGGAAAATTTACCCACCCTGAAGTGAG 50845

QY 181 CAAACTCGCCTCAGATTTAAATCTGGAGCGATGCTCTTTTGTCTCCGAGGTCTCTCAGTGA 240
DB 50844 CAAACTCGCCTCAGATTTAAATCTGGAGCGATGCTCTTTTGTCTCCGAGGTCTCTCAGTGA 50785

QY 241 AGAGACGATCTTCTCTGATCTTGGGAAAAAACAATGTACACAAGGCATTTATCTCAGCATG 300
DB 50784 AGAGACGATCTTCTCTGATCTTGGGAAAAAACAATGTACACAAGGCATTTATCTCAGCATG 50725

QY 301 CTGTGGCTTGGCAATGCTATTGTTTTGATGAGCGTATATATAGATTGAGGCGGTCTAT 360
DB 50724 CTGTGGCTTGGCAATGCTATTGTTTTGATGAGCGTATATATAGATTGAGGCGGTCTAT 50665

QY 361 CGTTTCGGAGAGCTGTTCTTCTGTAATCTTTTCTGTTATCTGGCAGCTCTACAGTATTTTGA 420
DB 50664 CGTTTCGGAGAGCTGTTCTTCTGTAATCTTTTCTGTTATCTGGCAGCTCTACAGTATTTTGA 50605

QY 421 TGCGCCACTCACCTTGTCTCAGGACTCGCTGGGATTTGTTCTTCTATGGGGATGGCGGTAGA 480
DB 50604 TGCGCCACTCACCTTGTCTCAGGACTCGCTGGGATTTGTTCTTCTATGGGGATGGCGGTAGA 50545

QY 481 TGCAAATGTTCTTGATTCGAAAGAAATCCGAGAGGAAATTTTTTATTTGTTCTCAAAGTCTTAA 540
DB 50544 TGCAAATGTTCTTGATTCGAAAGAAATCCGAGAGGAAATTTTTTATTTGTTCTCAAAGTCTTAA 50485

QY 541 AAAATCTGTAGAAAAGGATATACCAAGGCTTTTGGAGGCCATTTTGGATTTCTAACTTGAC 600
DB 50484 AAAATCTGTAGAAAAGGATATACCAAGGCTTTTGGAGGCCATTTTGGATTTCTAACTTGAC 50425

QY 601 TACAGTATGGCCTCAGCAGCTCTTTCTTCTCTAGATACAGGGCTATTAAGGGTTTGC 660
DB 50424 TACAGTATGGCCTCAGCAGCTCTTTCTTCTCTAGATACAGGGCTATTAAGGGTTTGC 50365

QY 661 TTTGACATGATTTTGAAGAAATTTTCTCTCAATGTTTACGGCTCTTTTTCATGACTAAAT 720
DB 50364 TTTGACATGATTTTGAAGAAATTTTCTCTCAATGTTTACGGCTCTTTTTCATGACTAAAT 50305

QY 721 TTTCTTCTCATGCTGGATGAATAGACCCACATACACAGTTGCTATATGATGAATAGTT 780
DB 50304 TTTCTTCTCATGCTGGATGAATAGACCCACATACACAGTTGCTATATGATGAATAGTT 50245

QY 781 CGTGGGATAAAGCATGATTTCTTCTGAGAGGATGCAAAAAATTTTGGGCTGTTTCTGGAAG 840
DB 50244 CGTGGGATAAAGCATGATTTCTTCTGAGAGGATGCAAAAAATTTTGGGCTGTTTCTGGAAG 50185

QY 841 TGTTTTTCTTTTAGTTGCGTGTCTCGGGTTTGGAGCCTTGAATTCGGTTTTTGGGAAT 900
|||||
```

Db 50184 TGTTTTCTTTAGGTTGCGTGTGCTCGGGTTTGAGACCTGAAATCCGTTTTGGAAAT 50125
 QY 901 GGAATTTAAAGAGGGATATGCTTTACCTTTAATCCAAAAGAGATGGCATCAGCGATGT 960
 Db 50124 GGAATTTAAAGAGGGATATGCTTTACCTTTAATCCAAAAGAGATGGCATCAGCGATGT 50065
 QY 961 TGTCTCAAAATCGTGGCAAAAGTTGTCATTAACAGAAAGCTGCTCTTTCTTTAAGA 1020
 Db 50064 TGTCTCAAAATCGTGGCAAAAGTTGTCATTAACAGAAAGCTGCTCTTTCTTTAAGA 50005
 QY 1021 CTTCCTGATTCAAACATTTGGATCTTCAGAAAGATCAAAATCTATTTTATGATTAAGC 1080
 Db 50004 CTTCCTGATTCAAACATTTGGATCTTCAGAAAGATCAAAATCTATTTTATGATTAAGC 49945
 QY 1081 TTTAAGCTATATCTAACAGATACGAGCTCTCTCTTAAATTAACATCATGAGCTGGC 1139
 Db 49944 TTTAAGCTATATCTAACAGATACGAGCTCTCTCTTAAATTAACATCATGAGCTGGC 49885
 QY 1140 GTTA-TTGTGGAGATTTGTCAGAAACAGGCTGATTTCTCTAGGAAACTCAAAACGA 1198
 Db 49884 GTTA-TTGTGGAGATTTGTCAGAAACAGGCTGATTTCTCTAGGAAACTCAAAACGA 49825
 QY 1199 AACGCAAAA-TTTTGTCAAAGATAGAGCAAACTATCGAAGAAAATGCGTTATCAGGC 1257
 Db 49824 AACGCAAAAATTTTGTCAAAGATAGAGCAAACTATCGAAGAAAATGCGTTATCAGGC 49765
 QY 1258 GACCATCGGGCTTTAAGAGCTTTGGCAATCATCTTGCTATATGAGTTGGCCTTTGA 1317
 Db 49764 GACCATCGGGCTTTAAGAGCTTTGGCAATCATCTTGCTATATGAGTTGGCCTTTGA 49705
 QY 1318 ATGGCAATATGCTTTCAGTGGCGATAGGCTTTAATCATGACCTTTTGGCTACCTGGC 1377
 Db 49704 ATGGCAATATGCTTTCAGTGGCGATAGGCTTTAATCATGACCTTTTGGCTACCTGGC 49645
 QY 1378 AGCTTTGTTATATGACATTTCTTTTGAAGAAAATCAAAATATGATTTGCAAGCCATGG 1437
 Db 49644 AGCTTTGTTATATGACATTTCTTTTGAAGAAAATCAAAATATGATTTGCAAGCCATGG 49585
 QY 1438 TGCTTTAATGACTGTATTTGGGATTCATTAACAATATCTTTGATTTGATCTGAT 1497
 Db 49584 TGCTTTAATGACTGTATTTGGGATTCATTAACAATATCTTTGATTTGATCTGAT 49525
 QY 1498 TCCGGAAGATCGCAAGGAACTGTTTACCCCTATGATGATTTTATGATGATGACCT 1557
 Db 49524 TCCGGAAGATCGCAAGGAACTGTTTACCCCTATGATGATTTTATGATGATGACCT 49465
 QY 1558 TCAAAAGACGTTACGCGCAAGGATATGACAAACAGTACCACTATCAGTTTGTAAAT 1617
 Db 49464 TCAAAAGACGTTACGCGCAAGGATATGACAAACAGTACCACTATCAGTTTGTAAAT 49405
 QY 1618 GCTTTTGTATTAAGGCGGCTCTCTGTCTTTAATTTTGTGATTTATGACCATAGGAT 1677
 Db 49404 GCTTTTGTATTAAGGCGGCTCTCTGTCTTTAATTTTGTGATTTATGACCATAGGAT 49345
 QY 1678 TCTTCTAGAACTTATATGCTCTCTTTAATATGACCAACCTGCTGTTTATGATGACCG 1737
 Db 49344 TCTTCTAGAACTTATATGCTCTCTTTAATATGACCAACCTGCTGTTTATGATGACCG 49285
 QY 1738 TTAAGAAAATCGCTCAAAATATAGTACGTTAACTTAACGTTAGCAATATAAA 1797
 Db 49284 TTAAGAAAATCGCTCAAAATATAGTACGTTAACTTAACGTTAGCAATATAAA 49225
 QY 1798 ATCTCTTTGGAGCTTATAGTCCCAAGGCGCTGTGATTTAAATTTATGCAAAATTCAG 1857
 Db 49224 ATCTCTTTGGAGCTTATAGTCCCAAGGCGCTGTGATTTAAATTTATGCAAAATTCAG 49165
 QY 1858 ATATATGC 1864
 Db 49164 ATATATGC 49158

RESULT 3
 AE002179

LOCUS AE002179 12198 bp DNA linear BCT 30-MAY-2000
 DEFINITION Chlamydomophila pneumoniae AR39, section 15 of 94 of the complete.
 ACCESSION AE002179
 VERSION AE002179.2 GI:8163380
 KEYWORDS
 SOURCE
 ORGANISM
 Chlamydomophila pneumoniae AR39.
 Bacteria; Chlamydiales; Chlamydiaceae; Chlamydomophila.
 REFERENCE
 1 (bases 1 to 12198)
 Read, T.D., Brunham, R.C., Shen, C., Gill, S.R., Heidelberg, J.F., White, O., Hickey, E.K., Peterson, J., Umayam, L.A., Uetzerback, T., Berry, K., Bass, S., Linher, K., Weidman, J., Khouri, H., Craven, B., Bowman, C., Dodson, R., Gwin, M., Nelson, W., DeBoy, R., Kolonay, J., McClarty, G., Salzberg, S.L., Eisen, J., and Fraser, C.M.
 Genome sequences of Chlamydia trachomatis Mohn and Chlamydia pneumoniae AR39
 Nucleic Acids Res. 28 (6), 1397-1406 (2000)
 2 (bases 1 to 12198)
 Read, T.D., Brunham, R.C., Shen, C., Gill, S.R., Heidelberg, J.F., White, O., Hickey, E.K., Peterson, J., Umayam, L.A., Uetzerback, T., Berry, K., Bass, S., Linher, K., Weidman, J., Khouri, H., Craven, B., Bowman, C., Dodson, R., Gwin, M., Nelson, W., DeBoy, R., Kolonay, J., McClarty, G., Salzberg, S.L., Eisen, J., and Fraser, C.M.
 Direct Submission
 Submitted (01-MAR-2000) The Institute for Genomic Research, 9712 Medical Center Dr., Rockville, MD 20850, USA
 On Jun 1, 2000 this sequence version replaced gi:7189108.
 FEATURES
 source
 1..12198
 /organism="Chlamydomophila pneumoniae AR39"
 /strain="AR39"
 /db_xref="taxon:115711"
 /note="synonym: Chlamydia pneumoniae AR39"
 145..1521
 /gene="CP0178"
 145..1521
 /gene="CP0178"
 /note="similar to GB:L26051.SP:P33986.PID:415662; identified by sequence similarity; putative"
 /codon_start=1
 /transl_table=11
 /product="UDP-N-acetylglucosamine
 1-carboxyvinyltransferase"
 /protein_id="AAF38052.1"
 /db_xref="GI:7189109"
 /translation="MOIAQVFCGRNGEVRKVSAGKNAATKLVASLSDOKTLRV
 PDIGVSLVELCKSLGAVSVSDKTEVLEIYETPEIICQRRVPTFSNVNRILILIGA
 LIGRCPEGVYPTVGGDAIGERTLTFHFEGLKQLOVOISSDSGGYAKAPRGLKGYT
 HLPYPSVATENLILAAIHAKRTVYIKVALEAELIDLVLFLOKAGADITTDNRID
 IFTGGLGSVDHTIIPDKIEAASFMAAVSGRFEVRNAKQELIPLKMRISGG
 FVSGSGIEFFQERPLVGVVLETVHPEGLFDWQOPEFVILSDQSSVIEHTVEN
 RGLYHGLQHWGABCOLPHQICSTACRAIENPFSVATHTATPMASHLVPIPLRA
 GRAYMALHIAAGGSIIEHTHILDRGTINWGRKSLGAKIQIDTMEBEELITPESK
 LAKRDASTL"
 complement(1475..3160)
 /gene="CP0179"
 complement(1475..3160)
 /gene="CP0179"
 /note="similar to PID:1001350.PID:1001346; identified by
 sequence similarity; putative"
 /codon_start=1
 /transl_table=11
 /product="arginyl-tRNA synthetase"
 /protein_id="GI:7189110"
 /db_xref="GI:7189110"
 /translation="NSTLISLISVCSQAIKAFPMLEMAPEITPSTKEHGYOCN
 DAMKLARVLRKAPRAIAEAI VALPEPSLLEIAGGINTFSFVFNQDLNHEKD
 ALKGFVQSPKKIIIDFSPNIANDMHVGRSTIIGDSLARIFSVYNDVLRNHI
 GMWGTAEMLITYLOENPCDYDLEDTLSLYKKAAYCFNDEEFKRSQGNVLAQAK
 DQALAIWEKICETSEKAPQKIYDILDIYVEKRSFYNPFLPEIIEDEKKGILLVS

gene
CDS
NDAKCVFHEAFSPFMVQKSGDGGVYATTDLAAMRYRIEBDHADKIIIVTDLGOSLHF
QLLEATAAGYQIFSHVGLVLDPOCKKLTGRSGENVKLRLELLDTAIEAEEA
LREHPELTAQIAPVIGINAIKYSDLSHSRTSDYFVSEKMLRFEQNTAFLLY
AVYRIQIGIRLIGISOLSGPPEIQBPABELLATLRRPEALESTIKELCHFLTD
YVNLTHKNGFRDSDHSDSPYAKSRFLCALAEQVATGMHLGLKTLERL
complement(3165..3803)
/gene="CP0180"
complement(3165..3803)
/gene="CP0180"
/note="1-acyl-sn-glycerol-3-phosphate acyltransferase,
putative; identified by match to PFAM protein family HMM
PF01553"
/codon_start=1
/transl_table=11
/product="1-acyl-sn-glycerol-3-phosphate acyltransferase,
putative"
/protein_id="AAF73638.1"
/db_xref="GI:8163381"
/translation="MIFRICKFFTWVAFSLFYKLVGVKQNFYKGPAAIIVNHNSEL
DPIALHMCVHECIYHLARSLFNIPWLKQMGCFVRODEGNSAAFKIASRLFNKRKK
LVYPEGARSPDQGPQKVGIGMMAAKSRVPIIPVYIRGTFFAFNRHQKIPHWKTI
TCVFGTPMYDDIIQNEIKNKETYQIITQTNKIAELKAWYESGCKGVP"
complement(3800..4450)
/gene="CP0181"
complement(3800..4450)
/gene="CP0181"
/note="similar to SP:P23863 PID:42839 GB:U00096
PID:1651431 PID:1651438; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="cytidylate kinase"
/protein_id="AAF38054.1"
/db_xref="GI:7189111"
/translation="MLITIDGPGCTGKSTAKALADHLHFNVCNTGKMYETLAVARLQ
SPWATLPLTFEPEPPSPFATGQOPLESFNGHLLSELITQEVANAASLSQLEPV
RAFMDLQRRYAQNGVCEFRDGMGSKVPNADLKIFLTSSPEVRAQRRLKDLPEGTL
SPEQLAELVKRDAADQRAHPLVPIENGIVDSSDLTTRQVLEKILALLFRNEL"
complement(4447..5373)
/gene="CP0182"
complement(4447..5373)
/gene="CP0182"
/note="similar to GB:ML1330 SP:P06466 PID:1208947
PID:145476 GB:U00096; identified by sequence similarity;
putative"
/transl_table=11
/product="phosphatidate cytidyltransferase, putative"
/protein_id="AAF38055.1"
/db_xref="GI:7189112"
/translation="MLNSNFKSKTGAYGDLFORVVVHSLVLTPLVLLLYSSLPFLTS
FALGFIATCGAVCTEYSSMAKAKHVPILSTFSAIGSFLLALSLISWGHSLPGF
FDALPWTLLVWVWSIFPRVKSTIGALQISGTLFSLVYGPDIRLFLVHVSFIHT
QBPVLGWMASFLIATGADIFGFFGKAGNKKIAPQISPNKTVGVFVAGCGATL
ISFIFFIQIPTRFASYPMPAILPLGLALGITGFFGDIIESIFKRDHALKSNKLKA
VGMMLDLSLLSTPIAYLPLITQSKERIG"
complement(5374..6159)
/gene="CP0183"
complement(5374..6159)
/gene="CP0183"
/note="similar to PID:1208946 GB:U00096 SP:Q47675
PID:1552751 PID:1786371; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="undecaprenyl pyrophosphate synthetase"
/protein_id="AAF38056.1"
/db_xref="GI:7189113"
/translation="MTVALINDPVDLSLATNNAESKPPSLQRLPNHVAIIMDGNRWY
KKHRECGHHTSGHYGAKVLNPLNAVLDLGLKVLTYTFSTENGRPKRQLQEIF
NIFYTDQKQLPUMENEICLRCIGDLSKLPKGIQTKINHVRNMTASFSRLELQVFN
YGGDELIVRAFKLHLDILNKKISSDLSLSTYDLSGLTDPDLLIRITGGMRVIS
NFLWQIAYTELYITDLPDFTPDQLFEAINVYQSRRRGGK"

gene
CDS
complement(6214..6363)
/gene="CP0184"
complement(6214..6363)
/genes="CP0184"
/note="hypothetical protein; identified by Glimmer2;
putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF38057.1"
/db_xref="GI:7189114"
/translation="MFKKPLYSEVTRLNFKLLETPICLNRIQPLPTGMAGEKSFVI
FVLGK"
6637..7737
/gene="CP0185"
6637..7737
/genes="CP0185"
/note="hypothetical protein; identified by Glimmer2;
putative"
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF38058.1"
/db_xref="GI:7189115"
/translation="MKPFRISISDLIPTQMWIRGGNVHYVYVNAQNLKKILGGVL
ACFGLALLGCAFAAGVQCOTIFPCIGLMILGLVLLGFAYLQYSKGRFRPLFRET
VPEKPINWLGLSLLQSKWKIRPGCYHPCQPOVEICEGSOEIVTKIFOKKSDRNTSI
FLIQEMDQIALROGIEKSSLSRKTFAIDPSVSSLLSEIOREEQOVLDPKVISWSSD
QASDRTHPKSAIYVNISSDAQEPQRCYDAYTKAFTVLDQIDQGNIVKHTIYVLT
PIUGVTHPKSEQENLKLLSQAAFYSQAQVAKRMREKQDSIRIKFIPTDTPSPS
LYFSPHSSHTPSVTPISLGFVGEQESYTF"
7917..12125
/gene="CP0186"
7917..12125
/genes="CP0186"
/note="similar to GP:3220156; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="secDF protein, putative"
/protein_id="AAF38059.1"
/db_xref="GI:7189116"
/translation="MKQVKRNFAIICVFAIYVYVLTCLYYAKPLDKKIDGNEAE
HIKSTTKAQQVRKDVIPVSAISSLHLRHQHPAIPDIVSVFRKGEQDAEDFI
GNLVHGEPNVPIKSARLVHVGYSREHDDHVIQVASSINTSLVSDRFSVYSSENEOE
MASSILQRYVSACTFPKQKDCSYPINETAKEQLLOLQYAKNLSGSEFVPSRLSAP
CQQFSNQDRFLAFSLSLSDAALDVEDQKLSVETLSQTACIRSLDCPFYIEG
LRUDCSSESLFFSSIBYCPKERKIFILTHSDLLAQRSLSKQRLDFSLRLAVEKQKL
SKNLTQVEDYNNFSGFQWMDKDTQGIILQGERLLQGIABHLTALTTLHRAAESCDL
IPENFPVFCRQPRESEAFGCYIFSPNTDCKHFGSKGVYILKGLRSTVAKYQOQGGKE
LOSFEKDLONLNFCSHTEAISWTLGEDQVLEIRHPLOQFLDVYMGEGFVIGKECAF
EVXDIQDRLATVNIQIKNRQSDLVVRHEQYRHAKCSMDLQERLSAPIPYONLFLNNK
LNNKFSRGNILRLGLDFVGGKQLLSFKDQKQLTDKEDILKVSDELCARLINKUG
VSBIEURREGDYIHLSPVGSSTISSEISLGTSTKXSPHVVNRFSSYSASRYEQRFID

Query Match 96.4%; Score 1797.4; DB 1; Length 12198;
Best Local Similarity 99.8%; Pred No. 0;
Matches 1831; Conservative 0; Mismatches 1; Indels 3; Gaps 3;
QY 1 ATGGACTTCGCATATTGTTCAGGAGGGGATCAGCGGCACCTGCTAATGACAATATTCTGC 60
DB 10364 ATGGACTTCGCATATTGTTCAGGAGGGGATCAGCGGCACCTGCTAATGACAATATTCTGC 10423
QY 61 AAACCGTGGATGGCGTATGCTGTAGTGATGACGGTTATATGTCAGCAGCCCTATTTT 120
DB 10424 AAACCGTGGATGGCGTATGCTGTAGTGATGACGGTTATATGTCAGCAGCCCTATTTT 10483
QY 121 AAACGTCCTATTGAAAATCATCGAGTGTCTCAGGAAATTTTACCCACCGTGAAGTCAG 180
DB 10484 AAACGTCCTATTGAAAATCATCGAGTGTCTCAGGAAATTTTACCCACCGTGAAGTCAG 10543
QY 181 CAAACTCGCCTCAGATTAAATCGGAGCGATGCTCTTTTGTCCCGAGGTTCTCAGTGA 240

Db 10544 CAAACGCTCAGATTTAAATCTGAGCAGATGCTTTGTCCCGAGGTTCTCAGGA 10603
 QY 241 AGAGACGATCTCTTCTGATCTTGGAAAAAACAATGACACAGGCAATTATCTCAGCATG 300
 Db 10604 AGAGACGATCTCTTCTGATCTTGGAAAAAACAATGACACAGGCAATTATCTCAGCATG 10663
 QY 301 CTGAGGCTGGCAATGCTTATGTTTGAAGAGCTATATATGATTTGAGAGCGTCAT 360
 Db 10664 CTGAGGCTGGCAATGCTTATGTTTGAAGAGCTATATATGATTTGAGAGCGTCAT 10723
 QY 361 CGCTTCGGAGCTGTTCTTCTGAATCTTTGCTTATCTGGGACGCTACAGTATTTGGA 420
 Db 10724 CGCTTCGGAGCTGTTCTTCTGAATCTTTGCTTATCTGGGACGCTACAGTATTTGGA 10783
 QY 421 TGGCCCACTACCTTTGAGAGACTGCTGGAGTTGTTCTGTATGAGGATGCGCGTGA 480
 Db 10784 TGGCCCACTACCTTTGAGAGACTGCTGGAGTTGTTCTGTATGAGGATGCGCGTGA 10843
 QY 481 TGCATAATGTTCTGTATCTGAAAGAAATCCGAGAGAAATTTTATATGCTCAAAGCTTAA 540
 Db 10844 TGCATAATGTTCTGTATCTGAAAGAAATCCGAGAGAAATTTTATATGCTCAAAGCTTAA 10903
 QY 541 AAAATCTGTAGAAAAAGATATACCAAGGCTTTTGGAGCCATTTTGTATCTTAATTTGAC 600
 Db 10904 AAAATCTGTAGAAAAAGATATACCAAGGCTTTTGGAGCCATTTTGTATCTTAATTTGAC 10963
 QY 601 TACAGTATGAGCTCAGCAGCTTTCTTCTCTAGATAACAGGCGCTATTTAAAGGTTTGC 660
 Db 10964 TACAGTATGAGCTCAGCAGCTTTCTTCTCTAGATAACAGGCGCTATTTAAAGGTTTGC 11023
 QY 661 TTTGACATGATTTTGAAGATTTTCTCTCAATGTTTACGGCTCTTTTCAAGCATTAAT 720
 Db 11024 TTTGACATGATTTTGAAGATTTTCTCTCAATGTTTACGGCTCTTTTCAAGCATTAAT 11083
 QY 721 TTTCTCAATGCTGATGATTAAGAACCAACATACAGTGTGATGATTAAGT 780
 Db 11084 TTTCTCAATGCTGATGATTAAGAACCAACATACAGTGTGATGATTAAGT 11143
 QY 781 CGTGGGATTAAGATGATTTCTTGAGAGATGCAAAAAACCTTTGGGCTGTTCTGAGG 840
 Db 11144 CGTGGGATTAAGATGATTTCTTGAGAGATGCAAAAAACCTTTGGGCTGTTCTGAGG 11203
 QY 841 TGTCTTCTTTTGAAGTGTGCTGCTGCTGGGTTTGGACCTGGAATCCGTTTGGGAT 900
 Db 11204 TGTCTTCTTTTGAAGTGTGCTGCTGCTGGGTTTGGACCTGGAATCCGTTTGGGAT 11263
 QY 901 GGATTTTAAAGAGGATGCTTACCTTAAATCCAAAGAGATGAGCATCAGGATGT 960
 Db 11264 GGATTTTAAAGAGGATGCTTACCTTAAATCCAAAGAGATGAGCATCAGGATGT 11323
 QY 961 TGTCTCAATGCTGAGCAAGTTGTGATTAACATACAGAGAGCTGCTCTTCTTGAAGA 1020
 Db 11324 TGTCTCAATGCTGAGCAAGTTGTGATTAACATACAGAGAGCTGCTCTTCTTGAAGA 11383
 QY 1021 CTTCGGATTTCAAACTTTGATGCTGCAAAAAAGATCAAAATCTTATGATTAAGC 1080
 Db 11384 CTTCGGATTTCAAACTTTGATGCTGCAAAAAAGATCAAAATCTTATGATTAAGC 11443
 QY 1081 TTTAAGCTATCT- AAGCAGATACAGAGCTCTCTCTTAAATTTAAGATCAATGAGCTGC 1139
 Db 11444 TTTAAGCTATCTAAGCAGATACAGAGCTCTCTCTTAAATTTAAGATCAATGAGCTGC 11503
 QY 1140 GTTA- TTGTGGATTTGTTGAGAAACAGGCTAGATTTCTTACGAACTCTTAAAGCA 1198
 Db 11504 GTTAAGCTGAGATTTGTTGAGAAACAGGCTAGATTTCTTACGAACTCTTAAAGCA 11563
 QY 1199 AAGCGAAA- TTTTGTGTAAGGATTAAGCAAACTATCGAAGAAATGCGGTATCAGGC 1257
 Db 11564 AAGCGAAAATTTTGTGTAAGGATTAAGCAAACTATCGAAGAAATGCGGTATCAGGC 11623
 QY 1258 GACCATGGGCTTTTGAAGAGCTTTGGCATCATCTTGTCTATGATGATTTGGCGTTGA 1317
 Db 11624 GACCATGGGCTTTTGAAGAGCTTTGGCATCATCTTGTCTATGATGATTTGGCGTTGA 11683

QY 1318 ATGGCAATATGCTTTCAAGTCCGCTATGCGCTTAAATTCATGACCTTTTGGCTACCTGTC 1377.
 Db 11684 ATGGCAATATGCTTTCAAGTCCGCTATGCGCTTAAATTCATGACCTTTTGGCTACCTGTC 11743
 QY 1378 AGTCTGTTATATGACATTTCTTTTGAAGAAATTCAAATGATTTGCAAGCAATGG 1437
 Db 11744 AGTCTGTTATATGACATTTCTTTTGAAGAAATTCAAATGATTTGCAAGCAATGG 11803
 QY 1438 TGCTTATATGATGCTATATGAGGATTTATCATTTAAACAATCTTTGATCATTTTGTGAT 1497
 Db 11804 TGCTTATATGATGCTATATGAGGATTTATCATTTAAACAATCTTTGATCATTTTGTGAT 11863
 QY 1498 TCGTGAAGATCGCCAGACCAAGCTGTTTACCCCTATGACATGTTTATGATGAGCCCT 1557
 Db 11864 TCGTGAAGATCGCCAGACCAAGCTGTTTACCCCTATGACATGTTTATGATGAGCCCT 11923
 QY 1558 TCATAAGAGCTTACAGCGGAGGTAATGACACAGCTACACTCTATCAGTTTGTAT 1617
 Db 11924 TCATAAGAGCTTACAGCGGAGGTAATGACACAGCTACACTCTATCAGTTTGTAT 11983
 QY 1618 GCTTTGTTTATAGCGGCTCTCTGCTTTTAAATTTTGAATTTATATGACATAGGAT 1677
 Db 11984 GCTTTGTTTATAGCGGCTCTCTGCTTTTAAATTTTGAATTTATATGACATAGGAT 12043
 QY 1678 TCTTCTAGACATTTATGCTGCTCTTTATATGACACCACTCTGTTGTTATGCTCCG 1737
 Db 12044 TCTTCTAGACATTTATGCTGCTCTTTATATGACACCACTCTGTTGTTATGCTCCG 12103
 QY 1738 TAAAGAAATCGCTCAAAATATAGTACGTTAACTTATATGATGATGATTAAT 1797
 Db 12104 TAAAGAAATCGCTCAAAATATAGTACGTTAACTTATATGATGATGATTAAT 12163
 QY 1798 ATCTCTTTGGGACTTTAGTCCCAAGGCCCTGT 1832
 Db 12164 ATCTCTTTGGGACTTTAGTCCCAAGGCCCTGT 12198

RESULT 4
 AE001318/c 9989 bp DNA linear BCT 30-OCT-2000
 LOCUS Chlamydia trachomatis section 45 of 87 of the complete genome.
 DEFINITION AE001318 AE001273
 ACCESSION AE001318.1 GI:3328875
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Chlamydia trachomatis.
 Chlamydia trachomatis
 Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
 REFERENCE
 1 (bases 1 to 9989)
 AUTHORS Stephens, R.S., Kaiman, S., Lammel, C.J., Fan, J., Marathe, R.,
 Aravind, L., Mitchell, W.P., Olinger, L., Tatunov, R.L., Zhao, Q.,
 Koonin, E.V. and Davis, R.W.
 TITLE Genome sequence of an obligate intracellular pathogen of humans:
 Chlamydia trachomatis
 JOURNAL Science 282 (5389), 754-759 (1998)
 MEDLINE 9900809
 PUBMED 9784136
 REFERENCE
 2 (bases 1 to 9989)
 AUTHORS Kaiman, S., Mitchell, W., Marathe, R., Lammel, C., Fan, J., Hyman, R.W.,
 Olinger, L., Grimwood, J., Davis, R.W. and Stephens, R.S.
 TITLE Comparative genomes of Chlamydia pneumoniae and C. trachomatis
 JOURNAL Nat. Genet. 21 (4), 385-389 (1999)
 MEDLINE 99206606
 PUBMED 10192388
 REFERENCE
 3 (bases 1 to 9989)
 AUTHORS Stephens, R.S., Kaiman, S., Lammel, C.J., Fan, J., Marathe, R.,
 Aravind, L., Mitchell, W.P., Olinger, L., Tatunov, R.L., Zhao, Q.,
 Koonin, E.V. and Davis, R.W.
 TITLE Direct Subinfection
 JOURNAL Submitted (20-MAY-1998) Program in Infectious Diseases, University
 of California, 235 Warren Hall, Berkeley, CA 94720-7360, USA
 FEATURES
 source 1..9989

```

/organism="Chlamydia trachomatis"
/strain="D/UW-3/Cx"
/db_xref="taxon:813"
complement(105..371)
/gene="omcA"
/note="CT444"
complement(105..371)
/gene="omcA"
/codon_start=1
/transl_table=11
/product="9kDa-Cysteine-Rich Lipoprotein"
/protein_id="AAC68043.1"
/db_xref="GI:3328876"
/transl_table="MKKTALLAALCSVVSLSSCCRIVDCFBPDCAPIQCSPECSEKKK
DVGCGNCSNGYVPACKPCGGDTHQDAKHGQARGIPVDGKCRQ"
708..932
/gene="CT444.1"
708..932
/gene="CT444.1"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAC68048.1"
/db_xref="GI:3522897"
/transl_table="MKRFELFIVATAALLTQGSYBPPEKREYILAQKCSPLFKHAQTL
DFSIAQAVFEMTFTTKAPPCVSSLELSQN"
complement(929..2449)
/gene="gltx"
/note="CT445"
complement(929..2449)
/gene="gltx"
/codon_start=1
/transl_table=11
/product="Glutamyl-tRNA Synthetase"
/protein_id="AAC68044.1"
/db_xref="GI:3328877"
/transl_table="MTVQNVVRVAPSPGDPHVGTAYMALFNEVPARKYNGOMILRI
EDIDTRSDRYANFISALKWCGIRWDEGPDVGGAYGPRQSERTEIYKXAEILLQ
TDCAYKCFATPQBLQENRAVASTGLYRGYDRYRLSPPEVEEQGPYTRILKY
PLTSGPVEQDCGCVVFWADVDQVLVKSDDGPTTFHFANVVDHLMGITHVLGRGE
WLSSTPHLLLYEAFGPEPPQFPHLLNPDGSKSKRNKPTSIIFYRDAGYKGEAF
MNFLLTLMGSMGEDEITYSMORLIEAFDPKRIGRSGAVPDIRKLDNMNKHVLAHNEGSP
ESILQELKGLWNEDEFLLKILPICOSRITTLADPVLGTSFFTFATFYQSKBELLPSSL
KQSQAAVMLYVLYLEKDLWEKDPFYQGSKWLAFAFVHHKKAVIDPLYVAITGAK
QGLPFLDSMELLGKARTRARLTVAQNLLGVSXKVKVQQQVDKALQDQPLEDIRLDF"
complement(2721..3272)
/gene="euc"
/note="CT446"
complement(2721..3272)
/gene="euc"
/codon_start=1
/transl_table=11
/product="CHLPS Euc Protein"
/protein_id="AAC68045.1"
/db_xref="GI:3328878"
/transl_table="MECLOQDTGVBAEQVQVQOQENAVPVTSQRVSITQAAKLHNV
RQAIYVAIKOKLKASKTRWEIDLOLEDYRNRYSRAKSTYQGLLFDNERGFYSV
QVAMSDVDPQIKYIVATRICAMKGERGSAWIVHSEVDRYRNDYLKKEARKGKSL
AAMREGFALGADLLADAENFIS"
complement(3677..5431)
/gene="recJ"
/note="CT447"
complement(3677..5431)
/gene="recJ"
/codon_start=1
/transl_table=11
/product="ssDNA Exonuclease"
/protein_id="AAC68046.1"
/db_xref="GI:3328879"
/transl_table="MEKDSLSLQKWDYPKHNEVFKTLKEFHLHPAQAQVLISRG
FQSIQETRDPLYSLSLHSTSLFDMEKAVARLLQAKANNEHVMYIGDDVDGIGTV
TLIVEFLQILGVKTSYCCSGTLPFOHGETASLISQMLQDGISLITVDDCGITAGKEVH
AINKQGIIDVITVDHHPMTGKLPHCVAMLPKLDKNPYPNKELTGVGVAFKLVCATYEE

```

```

LIQDASWKDKIDLLRFLDLVSLGTIADVGRSLGENRILVSYGICKIAKGRGLGLKL
CSLSGVDSKSEVSTLLNGIRITPKLNSGLKLAUSSQGVKLLSHDHPKNNIGTIVSELSAV
NOERQRIEAEVLDRVERILAAANPKTAQSAIVLASPNWHSRVPIIISARLARTYKPV
AIALQDGIKGLSGTIGFPLGLVRKESFPLSYGHDFAAGLMIKEDQVGGFRKK
FIHLVSSSLRKDDAMRTLSLDVMDPSRINRDLIASMELLEPGFKGNVSFVYTKAIO
VRYPKLACNHVKLYLNSCERNLEGTAGGQDKISLLKANMNIPLDIATVTLRMRSA
RGAIRLLIQDPRIQIPRL"
complement(5549..9751)
/gene="secD/secF"
/note="CT448"
complement(5549..9751)
/gene="secD/secF"
/note="fusion"
/codon_start=1
/transl_table=11
/product="Protein Export"
/protein_id="AAC68047.1"
/db_xref="GI:3328880"
/transl_table="MKRNFKNLSIILVFPALAHVLTCLYYSRPLNRKIDGREAR
KIERFTRQVSDARNDLTLURVAKILSTLKLGRDIKOPSPISGVNVHFFLAEDAALFI
DNVHGEPVPVKSRARLYVGSSEBFTGWQISGLVSLTSEKDSFVSYEDQVTY
SOBELNATNLFIESERPCEGYSLSMDMPIKVIYVTRNISCGLRLLPARKTQAPL
SRCLGTERDFTAPLARLEKALOHEEVKEAREELKQASLYLHLSRNLWNSISPKVGVNV
VDCSELSPCESMSLTADCKLAPHDAEVAANROQLVGNERLDLERLFAKEKOHIAIR
LMSVEGTGFAVRLRDEAASKIVLQHRVCQRIVEHLTALVLRNPITACTDLSSTE
NFPVYREPLDSDALGCFIPSPERSCRHFSKGSVIYFKGLRSIVAKYEHAGAEAL
LQHDQLNYACFIHTDASVSGEDRVLEIKEPQVIRIWIWEEFVQSGKASLEVRD
VRDLRAVNRRIEKTQHAELVRWDEQYRQACSNPRVLRRAAI PHKNIPEPILKLNIR
KYSRSHVLRFGTDFVGGQIIRIAPRDHGNLLTDKAGIDKYSDELVARLANKLVSEV
GWQREGDHLQSVPGGAAGISSADILGTSKMSHVUNEQFSRSPLEYQETFLDYLF
TARSLDECSQAINRLAGLPHDNGDAPANVRVAVEKREAGLSFSKLEGGASLD
TQYSMAIEKSEREQVNLPMIVFRNHALEGASLNIRPBFVAGEVGVNFGVXDKATF
SDRETVPQVQFHAMTSKFCQBGVSGTKNGLFGSGGRMAVAVLDGVVIDPVLNVPLK
DHASVGNFSYREVRHLAALDSGAMSFPEILSEVSWSPELGSSORVOQILSVLGL
VVLIVLVSYYFRFGVIAIVALLNLLIWSQYLDAPLTLSGLAGIILAMGAWDA
NVLVPERIPEEYLLTFSESVEAGYKAFSAIDENLTLTILASALLMLDTCPIKGF
ALTLLIGISSMFTALPMTKPFVFIWQKTRETQLHMMKFGIKKINFLKECRLLWV
SGVVLVGCVGGLGFGAWDFVGMDFGGYALTLDSDVCYNPEQMSVLKRRFQQIGL
SRDRVLRADSESEKVIYLSNALDRVQIEGAGEQSGDYHLARVLQVLSDSGSS
TSMVSDADSGWFKVSGOLSNKMTQAVIALFGALGILFYVSLRPEWRYAFSAICA
LMHDLATCAVLVAFHFQRIQIDQAIQALMTVLGYSNLNTLIIIPDRIDRERREKL
FTPMPELINDALQKTLGRVTMTATLSVLVILLFVGGGSIFNFAFMTVMTGILLGTLIS
SUIAPPLULFVWRKEEQNSLR"
BASE COUNT 3118 a 2399 c 1722 g 2750 t
ORIGIN
Query Match 41.5%; Score 773.6; DB 1; Length 9889;
Best Local Similarity 66.8%; Pred. No. 1.5e-161;
Matches 1170; Conservative 0; Mismatches 574; Indels 12; Gaps 4;
QY 2 TGGACTTCCGCATATTGTCAGAGGGGATCAGCGGCACTGCTAATGGACAATATTTCTGCA 61
DB 7318 TGGACATCAAGITTTGCCAAGAAGGGGTTAGTGGTACAAAAATGCGCTTTCTCTGGA 7259
QY 62 AACGTGGATGGCGTATGGCTGTAGTATTGACGGTTATATGGTCAGACGCCCTATTTTTA 121
DB 7258 GGGAGAGGGTGGCGTATGGCTGTATAGATGGCTACCGTATCAGCGATCCCGTGTG 7199
QY 122 AACGTCCCATTTGAAAATCATGCCAGTGTCTCAGGGAAATTTACCCACCGTGAAGTGCAG 181
DB 7198 AATGTTCCACTGAAAGATCATGTAGTGTTCCTGGAACATCTCTCATACAGAAAGTACAT 7139
QY 182 AAACCTCGCCTCAGATTTAAATCTGGAGCATGTCTTTTGTCCGAGGTTCTCAGTGAA 241
DB 7138 CGTTAGCTCAGATCTTAATCTGGAGCTATGTCTTCCATCCAGAAATTTTAAAGTAA 7079
QY 242 GAGACGATCTCTTGTGATCTTGGGAAAAAACAATGTACACAAGCATTTATCTCAGCATGC 301
DB 7078 GAAGTGTATCTCTGAGTTGGGAAGTTCTCAGCGTGTACAAGAAATCTCTCTGTTGTG 7019
QY 302 TGTGGCTTGGCAATGCTTATGTTTGTATGAGCGTATATATAGATTTGGAGGCGTCATC 361
DB 7018 TTGGGACTTGTAGTGTCTCATCTAATAAGTGTGTACTACAGATTTGGAGGCTGTATT 6959

```

QY 362 GCTTGGAGGCTGTTCTTGAATCTTTGCTTATCTGGGAGCTTACAGTATTTGGAT 421
 DB 6958 GCTTCTATAGGCGTGTGCTGCTAAATCTATTTGCTCAATTTGGGCTTTATGCAATATTTGGAT 6899
 QY 422 GCGGCACTCACCTTGTGAGGAGCTGCTGGAGTGTCTTCTGATGGGAGTGGCCGTAGAT 481
 DB 6898 GCTCCACTTACTCTATCCGGGTTAGCTGTATTTATCTTACAGAAATGGAGTGGCTGTGAT 6839
 QY 482 GCAATGTTCTTGTATTCGAAAGATCCGAGAGAAATTTTATGTCTCAAAATCTTAA 541
 DB 6838 GCTATGTTCTAGCTTTGAAAGATCCGAGAGAAATCTTTGACTAGAAAGCTTGCG 6779
 QY 542 AAATCTGAGAAAAAGATATACAGAGCTTTTGAGGCAATTTTGATTTCACTTACT 601
 DB 6778 GAGCTGTAGAGAGGTTATATAAAGGCTTCAGTGCATTTTGTGACTTAAATTTTACT 6719
 QY 602 ACAGTATTTGGCTCAGGACTCTTTTCTTCCAGTATACAGGAGCTTAAAGGGTTGCT 661
 DB 6718 ACTATTTTGTAGCTTGTGCTGCTGATGTTAGTATACAGAGCTTATCAAAAGTTTGGCG 6659
 QY 662 TTGACATTTGATTTAGGAATTTTCTTCAATGTTTACGGCTTTTTCATGACTAAATTT 721
 DB 6658 CTCATTTGATCATCGGATTTTCTCATCCATGTTTCAAGCCGCTTTATGACGAAATTT 6599
 QY 722 TTTCTCATGCTGTGAGATATAGACCAATACACATGCTGATATGATATATGATTC 781
 DB 6598 TTTCTTGTGCTATTTGGTGACAAAAAGAGAGACCCCAATTTGCAATATGATATATAGTTT 6539
 QY 782 GTGGGATTAAGCATGATTTCTTGAGAGATGACAAAAAATTTGGGCTGTTCTGAGAGT 841
 DB 6538 ATTGGATCAAGCATTAATCTTTTAAAGAAATGTAAGCTTTGTGGCTATGTTTCTGAGATA 6479
 QY 842 GTTTTCTTTTATAGTGTGCTGCTCCGGTTTGGAGCTGGAATTCGTTTGGGAATG 901
 DB 6478 GTTCTGTTTATAGTGTGTGAGATATAGGTTTGGTGTGCTGGGAATTCGTTTGGGATG 6419
 QY 902 GATTTTAAAGAGGATATGCTTTTATCTTTATCCAAAAAGCATGGATCAGCATGTTT 961
 DB 6418 GATTTTAAAGGATGATGATGACATGCTTTAGTTCGATGCTGTGATGATGATGAC--ATCCA 6362
 QY 962 GCTCAAAAGCTGGCAAGTGTGATTAACATCAGAGAGCTGGCTTCTTCTTACAGAG 1021
 DB 6361 GAGCAGATGTGTCTGATTTAGAAAGGCTTCCAAATTCGATTTCTTCTGAGAT 6302
 QY 1022 TTCCGATTTCAAAATTTGATCTTTCAGAAAAAGATCAAAATCTATTTAGTATTAAGCT 1081
 DB 6301 TATGCTGTGCTGAGAGATATGTTCTGAAAAAGTCAAAATATATTTATCTCAGAAATGCC 6242
 QY 1082 TTAAGCTATACTAGCAGATACAGACCTCT-----CTCCTAAATTAAGATCATGAG 1134
 DB 6241 TTGATAGATTTGAGCAATATAGAGAGCTGTGATGAGAACGAAAGATCGGATTTATCAT 6182
 QY 1135 CTGGGCTTA-TTGCGGAGTGTGTGTCAGAAAAGGCTGATATTTCTACGGAATCTCTA 1193
 DB 6181 CTGTGCTGAGTCTTCAAGTATGTTCTGATTTCCGAAAGTTCTACTACTTCTATGATATTT 6122
 QY 1194 AACGAAACGCAAAATTT-TGCTCAAAAGTAAAGCAGCAAACTATTCAGAAAAATGCGTTAT 1252
 DB 6121 GATGCTTCTCAGAGTATGCTGTTTAAAGTAAAGTGAACGCTTTCAATAAAAATGCGTACA 6062
 QY 1253 CAGGCGACATCGGGCTTTTAAAGAGCTTTGGCATCATCTGCTCTATGATGATTTGGCG 1312
 DB 6061 CAGGCTGTCATCGCAATATTTGAGAGCTTTAGGAATTAATCTTATTTATGTTAGCTTGGC 6002
 QY 1313 TTTGAAATGCAATATGCTTTGATGCTGCTGATGCTGCTTAAATTCATGACCTTTGGCTACC 1372
 DB 6001 TTTGATGAGATATGCTTCAAGTGTGCTTTTAAAGCAGACTTATTAAGCACT 5942
 QY 1373 TGTGAGCTTGTGTTATAGACATTTTCTTTTGAAGAAATCAATATGATTTGCAAGCC 1432
 DB 5941 TGTGCTGTGCTAGTGTGCTTACATTTCTTTTTCAGAGAAATACAGATTTGATTTGCAACA 5882

QY 1433 ATTGAGCTTTAATGACTGATTTAGGGATTTTCATTAACAATATCTTGATCTTTTATGAT 1492
 DB 5881 ATAGGCGATTTGATGATGACAGTGTAGGGTACTCTTTGAAACAATACGTTGATATTTTGGAT 5822
 QY 1493 CGTATTTGTGAGAGATCGCCAAAGGAACTGTGTTTAACTCCCTATGATGATTTTATGATAT 1552
 DB 5821 CGTATTTGAGAGATCGGCGCAAAAACTGTTCACCCCAATGCGGATTTGATCAATGAT 5762
 QY 1553 GCCCTTCAAAAGAGCTTACAGCCGCAAGGATATACAAACAGCTCAACTCATAGTTTGG 1612
 DB 5761 GCATTTGAGAGACATTTAGAGCTTACAGTATATACAGGCAACAATCATATCTGATG 5702
 QY 1613 TTAATGCTTTTGTTTATAGGCGGCTCCTGCTCTTAAATTTTGCATTTATATGACCA 1672
 DB 5701 GTCAATTTTGTATTTCCGTGGGAGAGGTTCTATTTTCAATTTTGCATTTATGACAGTG 5642
 QY 1673 GGGATCTTCTAGGAACCTTTATGCTCTCTTATATGACACACCTCTGTGTTGTTATG 1732
 DB 5641 GGGATCTTGTAGGAACGCTATCTTATGATATGATAGCTCTCGCTTCTTATTCATG 5582
 QY 1733 GTCCGTAAAGAAATC 1748
 DB 5581 GTGCTAAGAGAGAAC 5566

RESULT 5
 AE002341/c
 LOCUS
 DEFINITION
 Chlamydia muridarum, section 69 of 85 of the complete genome.
 ACCESSION
 AE002341 AE002160
 VERSION
 AE002341.2 GI:8163305
 KEYWORDS
 SOURCE
 Chlamydia muridarum.
 ORGANISM
 Chlamydia muridarum
 Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
 REFERENCE
 1 (bases 1 to 12442)
 Read,T.D., Brunham,R., Shen,C., Gill,S.R., Heidelberg,J.F.,
 White,O., Hickey,E.K., Peterson,J., Unayam,L.A., Uetreback,T.,
 Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B.,
 Bowman,C., Dodson,R., Gwinn,M., Nelson,W., DeBoy,R., Kolonay,J.,
 McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.
 Genome sequences of Chlamydia trachomatis Mopn and Chlamydia
 pneumoniae AR39
 Nucleic Acids Res. 28 (6), 1397-1406 (2000)
 JOURNAL
 MEDLINE
 20150255
 PUBMED
 10684935
 REFERENCE
 2 (bases 1 to 12442)
 Read,T.D., Brunham,R., Shen,C., Gill,S.R., Heidelberg,J.F.,
 White,O., Hickey,E.K., Peterson,J., Unayam,L.A., Uetreback,T.,
 Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B.,
 Bowman,C., Dodson,R., Gwinn,M., Nelson,W., DeBoy,R., Kolonay,J.,
 McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.
 Direct Submission
 TITLE
 Submitted (01-MAR-2000) The Institute for Genomic Research, 9712
 Medical Center Dr, Rockville, MD 20850, USA
 COMMENT
 On Jun 1, 2000 this sequence version replaced gi:7190754.
 FEATURES
 source
 1..12442
 /organism="Chlamydia muridarum"
 /db_xref="taxon:83560"
 /note="synonym: Chlamydia trachomatis Mopn"
 complement(86..544)
 /gene="TC0726"
 /gene="TC0726"
 complement(86..544)
 /gene="TC0726"
 /note="similar to SP:P18587; identified by sequence
 similarity; putative"
 /codon_start=1
 /transl_table=11
 /product="15 kDa cysteine-rich outer membrane protein"
 /protein_id="AF39536.1"
 /db_xref="GI:7190755"
 /translation="MSTTPIVSGVTSQNNSENVSNNARSLTLKERASKLSSTAFKY

GLAVGIFLVLSIVLFIPLPATASNPYLAIPAILGCVNICIGILSMWKGSCSEAK
WKLCKNVKTSIEDLDGNNKNIPTDNLNRVEDIVITLSSRRNSVA"
/complement (727. .2391)
/gene="TC0727"
complement (727. .2391)
/gene="TC0727"
/notes="similar to ; identified by sequence similarity;
putative"
/codon_start=1
/transl_table=11
/product="60 kDa outer membrane protein"
/protein_id="AAF39537.1"
/db_xref="GI:7190756"
/translation="MRIGDPMKLIIRRAVTFIPATVSASLPASGLVETSMASLSITNV
ISLADTKAKETTHQKDKKQNHNTSVVRKEVTAARDTKAVEPRQDSCFGKMYTV
KANDRNEVTVQSPYATYGSPIETAIKGRDCVDVITIQOLPEAFSVSDPAT
TPVADKLWIKDRIGQGEKSKITVWPKLKEGCCPTAAATVCACPEIRSVTKCGQPAI
CVQGEPSACLPCTYRINNVNNGTATARNVVENPVDGVAHASGQRLVLTGLD
MORGEORTIVCEPLKGRVTNATVSYCGGHKNTASVTIVINEPCVQVNIAGDWS
YVCKPBYIVSVNPGDLVRDVIEDTLSPGIVTVEAQAQISCNKLWILKELNPG
ESIQYKLVRAQTPGQFTNNVVKSCSDCGICTSCAEATYWGKVAATHMVCVVDTCDP
ICGVNTVYICVTNRGSADNTNVLILKFSKELQPTISFGPTKGTITNTVTFDSLP
RLGSKTFEVSVLKAVSAGDAGEAILSSDLTLPVPSDTEHTHY"
/complement (2539. .2805)
/gene="TC0728"
complement (2539. .2805)
/gene="TC0728"
/notes="similar to ; identified by sequence similarity;
putative"
/codon_start=1
/transl_table=11
/product="cysteine rich outer membrane protein 3"
/protein_id="AAF39538.1"
/db_xref="GI:7190757"
/translation="MKKTALLAALCSVLSLSSCRIVDCFCFDPCEPIKSPCESKRR
EVNGGCSNGYVPSCKPCGGELDHETKQGPARGIQADGRQ"
/complement (3357. .3357)
/gene="TC0729"
3357. .3357
/gene="TC0729"
/notes="conserved hypothetical protein; identified by
Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF39539.1"
/db_xref="GI:7190758"
/translation="MKRFFLMATAALLTQGCSEYPPKREYILAQRCSPLFKAQAI
NFSLAQVFKFTVTVQPSLNKLKLSQN"
/complement (3354. .4874)
/gene="TC0730"
complement (3354. .4874)
/gene="TC0730"
/notes="similar to PID:1783380; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="glutamyl-tRNA synthetase"
/protein_id="AAF39540.1"
/db_xref="GI:7190759"
/translation="MTQFNVRVRVAPSTGDPHVGTAYMALFNEVFARKFNGKMLIRI
EDTDQSRSDYDQENRFAISALKWCGIRWDEGPDIGPGPYRQSERTEIYKKAIEILLQ
TDYAKCPQEOLEKRAVASTLGYRGYDRRYRVLSEVRQEEQGPYTRILKV
PLTGESVFQCKGRVFPWADVDQVLKSDGFPPTVHPANVDDHLMGITHVLRGEE
WLASTPQHLLYEAFGEWEPQPFHMLLNPDLGSKLSKKNPTISFYRDAYKKEAF
MNFLLTNGYMEGDEEYISQRLIEAFDPKIRGSAVFDIRKLDNMKNHLYNEHGP
ESLLEKULWMDNEFLKLPQCRTITTLADFIRLTSFFMAIPEYSKEDLPSLL
GHQAAVMLYSVLEKOLWEKDFYQGSKLAEAFVQHHKAVIPLLYVTITGEK
QGILPDSMELLKARTRLTAQNLLGGVPKKLQQIDKLDQDQPLEBIRLDF"
/complement (5124. .5681)
/gene="TC0731"
complement (5124. .5681)
/gene="TC0731"

/notes="similar to GB:L13598 SP:Q06566 PID:289835;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="euo protein"
/protein_id="AAF39541.1"
/db_xref="GI:7190760"
/translation="MECLOQDTWAEVQEQEEQEEENAVSANAQRVSTOAAKLHNH
TRQAIYVAIKQKKKASYATRIKAMKGERGSAWVIHVSEVDYRNDLYLKEABRKGQN
VQVAMLVNPEQIKYATRIKAMKGERGSAWVIHVSEVDYRNDLYLKEABRKGQN
LSAMREGFTGLGASELLADSEDFVS"
/complement (6143. .7891)
/gene="TC0732"
complement (6143. .7891)
/gene="TC0732"
/notes="similar to GB:M54884 SP:P21893 PID:147550
PID:887842 GB:U00096; identified by sequence similarity;
putative"
/codon_start=1
/transl_table=11
/product="single-stranded-DNA-specific exonuclease RecJ"
/protein_id="AAF39542.1"
/db_xref="GI:7190761"
/translation="MEKDKLSALGQWTPKNEEFLLKILKELYLHPVIAQVLVSRG
FQSVQIRDFLYPQFSLPASPFLDMERKAVARLLKSKENNEHVMYIGDGDVGTIGV
TLLEFLQMLGVKTSYCCSGTLFQKHGEITSLISQMLQDGVSLITVDCGTTAGREVR
SINKQIDIVITDHTPTKPCIAIATLPKLDKNSYPNKELTGVGAPKLVATAYEE
LIKQDASWKDKIDLLHFLDLVSLGTIADVGRSGENRILVSYGIKETTKGRLLGKLL
CSLQSGKSDISSTNLGIRITPKLNSLGRLASQGVKLLLSQDPRVTIVSELAIV
NQERIEAEVLRDVERILANPKLTAQSAIVLASPYWHSRVPISARLARTNKPV
AIALQDQIGKSLRTIGAPFLDGLVRKCESFSLSYGGHDFAAAGLWIKHEQIEGFRKK
FIHLVNSLRKDEATLPLDGMDFSQISQDLISSMELLEPEFGKGNPSPVFTTATQ
VRYPKLLSGNHVKLYLNSGERNLEGTAFGGDKISLLKANNMPLDIATYTLRWKRSV
HGAVRLIIODERIOAP"
/complement (8009. .12211)
/gene="TC0733"
complement (8009. .12211)
/gene="TC0733"
/notes="similar to GB:AL009126; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="secDF protein, putative"
/protein_id="AAF39543.1"
/db_xref="GI:7190762"
/translation="MKRNFKNLSVIIFFALALYHVLPTCLVYSRPLNKKIKGREAH
KIIERLTQVADARNDLTVRKRLSTLKLRSVKPHLSIPGVVNVHFLAEDAAPVI
DNVHGEVPTKARLYVIGSELGTETIVQVSGSLVTSLTBEEDSFVSDEKDLVW
SOELENIATNLFVNVECDHGYSLWDMPIEKVAHARGICSGRLRLPTEKQAPL
SRPLGTEDQFVAPLSRVENALHEDYKEYREELNEAAYLLHSNRLNWNENSPRVGV
VDCSELSPCFSSMSLTGDGKLLPHFDADVIKQQLVGBERDLRLFAVEKQHIATR
LNRSVEETSGPAFLRDEAAGKIVLQGRVQRIVEHLTALVLRPIAETCDLSTE
NPVYSREPMENDTLGCFIPSPSSCRHFSKGVIVFKGLRSIILACESSGQBEAGL
LQODLQNLVACFIHTDAVSGEDRVLEIKEPLQIVIQWGEFVQSPGEASLEUVL
VDRLAVERNRIEKAHAELVRMDQYRQCSMNQVRLRAAI PHKRVFVFNKLVNR
KYSRGEHLRFQGTDFVGGIRLAFRDHQNLTVDKEGIDKVSDELYARNKLGSEV
GMOREGDHIQVSPVGAANLSDILCTSKMSHVNVNEOFSSRSPRYEVOTFDLYWF
TARLSDKSPKDIINHLAATI FHEHQGI PANVPAAI GKLEAGLNSKEHEMSACID
TQYSMAIVKSGSGEQVNLMI VFRNHALGASLKNIRPEFVAGEGVNFGVQDKAVS
LUSKTPVOOFHAWTSKFCOEVNGTKNSQFSGGRMVAVLVDGVVSDPVLNVPUK
DHASVNSFYSREVNRLASDLKSGMSFTPEILSEEVSPELGSLORLOQILSVFLGL
VVLIAWVSYYKFGVGVASVALLNLLI WASQYLDAPLTLSGLAGILLAMGWADA
NVLVFPIREBYLLSRSLSESBAVYKAFSAI FDSNLTTLASALLMLCKDTGPIKGF
ALTLIIGIPSSMFTSLFMTKFPFLIWKTRTQLHMMNKFIGVKNFLKECKRLWV
SGAVVLGICIGIGFSAWDSILNDMPKGGVALTLDSDVCAYNPQMSALKKRFREGL
SRDXYRADSQSGKVIYIYNALARVQAKGSEYEQNGADHHLAQVLQVLSDDSA
TSPAPAESRSGNFWKVGSLQSNKMTQAIALMGALGIILLYVSLRFEWRYAFSAICA
LMHDLIATCSVLVAFHRLRIQIDLQAVGALMTVLGYSLNNTLIIIFDRIRERREKL
FTPMPTLINDALQKTLGRTVMTTATLSVLVILLFVGGSGIFNFATIMTIVILLGLTILS
SLYIAELLLFMVKEEQNSLO"
3968 a 2882 c 2074 g 3518 t
BASE COUNT
ORIGIN

Query Match 40.8%; Score 760.8; DB 1; Length 12442;
 Best Local Similarity 66.2%; Pred. No. 1e-158; Indels 12; Gaps 4;
 Matches 1162; Conservative 0; Mismatches 582;

QY 2 TGGACCTCCGATATTTGTCAGAGAGGATCAGCGCATCTGTAAATGACATATTTCTGCA 61
 DB 9778 TGGACATCTAAGTTTGTTCAGAGGGGGTTATGGAACAAAATATGCACTTCTGGA 9719

QY 62 AACGGTAGGCGGTATGCTGTATGATGATGACGGTTATATGTCAGACGCCCTATTTTA 121
 DB 9718 GGAAGAGGTGGCCATGCGCTGTGTGTTTATGATGGTATGATATGATGATCCCGTGT 9659

QY 122 AACCTCCATGAAAAATGTCAGAGGTCTCAGGAAATTTACCCACCGTAAGTGC 181
 DB 9658 AATGTTCGCTCAAGATGCAAGAGTGTCTGGAATTTTGTACAGGGAAGTCAT 9599

QY 182 AAACCTGCGATATTTAAATCTGAGAGCATCTTTTGTTCGAGGTCTCAGTCAA 241
 DB 9598 CGTTAGCTTCTGATCTTAATCTGGGCTATGTCCTTTATCCCTGAATTTTAAGTGA 9539

QY 242 GAGACGATCTTCTGATCTTGGGAAAAAATGATACAGAGCATTTATCTCAGCATGC 301
 DB 9538 GAAGTATTTCTCAGAGTATGAGACATCTCAGAGTTGCAAGGATCTCTGTATTT 9479

QY 302 TGTGCTTGGCAATGCTTATTTGTTGATGAGCGTATTTATGATTTGGAGGGTCAATC 361
 DB 9478 TTAGAGCTAGTGTCTCATATGCAATTAATGATGATTAATTAATGAGTGGGGTGTCT 9419

QY 362 GCTTCGAGAGCTGTCTCTGCAATCTTTGCTTATCTGGGAGCTCTCAGATATTTGGAT 421
 DB 9418 GCTTCGAGAGCTGTCTCTCTTAATTTACTTACTTATCTGGGATCTATGCAATTTTGA 9359

QY 422 GCGCATCTACCTTGTGCAAGCTGCTGGATGTTGTTCTGTATGAGGATGCGGTATGAT 481
 DB 9358 GCTCCTCTTATCTATCAGAGCTAGCTGGAATTAATCTAGCAATGGGATGCGGTATGAT 9299

QY 482 GCAATGTTCTTGTATTTGGAAGAAATCGAGAGAAATTTTATGTTCTCAAGTCTTAA 541
 DB 9298 GCTAAGCTGCTGTTTGAAGAAATGCAAGAAATTAATTAAGTAGAAGTTATCA 9239

QY 542 AAATCTGAGAAAAAGATATACCAAGCTTTTGGAGCATTTTGTATCTTCACTTGA 601
 DB 9238 GAGCTGTAGAGCTGTGTTATGAGAGCTTTTATGCAATTTTGTATCTTAATTAACA 9179

QY 602 ACAGTATTTGCTCAGACATCTTCTTCTCTGATATACAGGCGCTATTAAGGTTGCT 661
 DB 9178 ACTATTTTGGCTGCGCATGCTCTCTCATGTTGATACAGACCAATTAAGGTTGCT 9119

QY 662 TTGACATTTATTTAGAAATTTCTCTCAATGTTTACGGCTTTTTCATGCTAAATTT 721
 DB 9118 TTGACCTGATCATTTGAAATTTTCTCTATGTTCAATCTTTGTTATGACGAAGTTC 9059

QY 722 TTCTTCATGCTGTGATGAATTAAGACCAACATACAGAGTTGATGATGAATTAAGTTC 781
 DB 9058 TTTTCTTGAATTTGATACAAAAAGAGAGACCAATTAACATTAAGTGAAGTTC 8999

QY 782 GTGGGATTAAGCATGATTTCTTGAAGAGATGCAAAAACTTTGGGCTGTTCTGGAAT 841
 DB 8998 ATAGGGGTCAACATTAATTTCTTGAAGAAATGTAAGAGCTTTGGGTTGTTCCGGGCT 8939

QY 842 GTTTTCTTTTATGTTGCTGCTCTCGGTTTGAAGCTTGAATTCGTTTGGGAATG 901
 DB 8938 GTAGTGTGTTTATGTTGATGAGGCTTGAAGCTTTGGTGTCTGGGATCTCATTTTGAATG 8879

QY 902 GATTTTAAAGAGGATGATGCTTTTATCTTTATCCAAAAGCATGATGATGATGATG 961
 DB 8878 GATTTTCAAGAGAGATACCGCTGATCTTGAATTTGATGTTTGTCTTAAT---ATCCA 8822

QY 962 GCTCAATGCTGCGCAAAAGTGTGATTAACCTACAGAGCTGTCTTTCTTATGAGAC 1021
 DB 8821 GATCAGATGTGTCTGTTTAAAAAAGCTTCGCGAAATTTGGGTTGTCTTCTGCTGAT 8762

QY 1022 TTCGATATCAACATTTGATCTTCAAAAAAGATCAAAATCTATTTTATGATTAAGCT 1081

DB 8761 TATGCTGTGCTTAAGCTATGATCTCTGAAAAAGTTTAAGATCTATTTTCTCAAAATGCT 8702:
 QY 1082 TTAAGCTATTAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1134
 DB 8701 TTAAGCTATTAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 8642
 QY 1135 CTGCGCTGTA-TTGTGGATTTGTTGTCAGAAACAGGCTGATTTCTCTACGGAATCTTA 1193
 DB 8641 CTAGCTCAAGTTTGAAGATTAATCCGATTTAGATGATGATGATGATGATGATGATG 8582

QY 1194 AACGAAACGCAAA-ATTTTGTCAAGATGATGATGATGATGATGATGATGATGATGATG 1252
 DB 8581 GAATCTCTGAGAGAGTGTGTTAAGTATGATGATGATGATGATGATGATGATGATGATG 8522

QY 1253 CAGGCGCATGCGGCTTTTGAAGCTTTGCAATCATCTGCTATGATGATGATGATGATG 1312
 DB 8521 CAGGCAATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 8462

QY 1313 TTTGAATGCAATATGCTTTTCAAGTGTGCTTATGATGATGATGATGATGATGATGATG 1372
 DB 8461 TTTGAATGCAATATGCTTTTCAAGTGTGCTTATGATGATGATGATGATGATGATGATG 8402

QY 1373 TGTGAGTCTGTTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1432
 DB 8401 TGTGCTGTATTAAGTGTGCTGATGATGATGATGATGATGATGATGATGATGATGATG 8342

QY 1433 ATGTGCTTATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1492
 DB 8341 GTGGGCGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 8282

QY 1493 CGTATTCGTAAGATGCGCAAGGAACTGTTAACCCTATGATGATGATGATGATGATGATG 1552
 DB 8281 CGTATTCGTAAGATGCGCAAGGAACTGTTAACCCTATGATGATGATGATGATGATGATG 8222

QY 1553 GCCCTCAAAAGAGCTTACAGCCGCAAGTATGATGATGATGATGATGATGATGATGATG 1612
 DB 8221 GCTTGTGAGAAAGCTCTGAGAGCTGATGATGATGATGATGATGATGATGATGATGATG 8162

QY 1613 TTAATGCTTTTGTATTAAGGCGCTCTCTGCTTTAATTTGATGATGATGATGATG 1672
 DB 8161 GTCATTTTGTATTTGTTGAGAGTGTGCTCTTATTTCAATTTTGTGTTATTAAGCGGTA 8102

QY 1673 GGGATTTCTTATGAACTTATGCTCTTATTAATGATGATGATGATGATGATGATGATG 1732
 DB 8101 GGGATTTTGTATGAACTTATGCTCTTATTAATGATGATGATGATGATGATGATGATG 8042

QY 1733 GTCGTAAGAAATC 1748
 DB 8041 GTGCTAAGAGAGAC 8026

RESULT 6
 AE010580 11527 bp DNA linear BCT 25-MAR-2002
 LOCUS AE010580
 DEFINITION Fusobacterium nucleatum subsp. nucleatum ATCC 25586, section 122 of 197 of the complete genome.
 ACCESSION AE010580 AE009951
 VERSION AE010580.1 GI:19714220
 KEYWORDS
 SOURCE
 ORGANISM
 REFERENCES
 AUTHORS
 Kapral, V., Anderson, I., Ivanova, N., Reznik, G., Los, T., Lytkidis, A., Bhattacharya, A., Bartman, A., Gardner, W., Grechkin, G., Zhu, L., Vasileva, O., Chu, L., Kogan, Y., Chaga, O., Goldsman, E., Bernal, A., Larsen, N., D'Souza, M., Walunas, T., Pusch, G., Haselkorn, R., Fomstein, M., Kyriades, N. and Overbeek, R.
 Genome sequence and analysis of the oral bacterium Fusobacterium nucleatum strain ATCC 25586
 JOURNAL U. Bacteriol. 184 (7), 2005-2018 (2002)
 MEDLINE 2186394

11889109
 2 (bases 1 to 11527)
 Kapatal,V., Anderson,I., Ivanova,N., Reznik,G., Los,T.,
 Lykidiis,A., Bhattacharyya,A., Bartman,A., Gardner,W., Grechkin,G.,
 Zhu,L., Chu,L., Kogan,Y., Chaga,O., Goltzman,E., Bernal,A.,
 Larsen,N., D'Souza,M., Walunas,T., Pusch,G.D., Haselkorn,R.,
 Fonestein,M., Kyripides,N. and Overbeek,R.
 Direct Submission
 Submitted (13-FEB-2002) Integrated Genomics, 2201 W. Campbell Park
 Drive, Chicago, IL 60612, USA
 Location/Qualifiers
 1. 11527
 /organism="Fusobacterium nucleatum subsp. nucleatum ATCC
 25386"
 /strain="ATCC 25386"
 /sub_species="nucleatum"
 /db_xref="ATCC:25386"
 /db_xref="taxon:190304"
 414..2345
 /gene="FN0694"
 414..2345
 /gene="FN0694"
 /codon_start=1
 /transl_table=11
 /product="S-layer protein"
 /protein_id="AAL94890.1"
 /db_xref="GI:19714221"
 /translation="MNETTQSKLVTDVKTQYQYVDPVSVYIEKEPNSDIINIYGKIK
 GNPYSLADKGEYVNDKIRIFGNVDITSTKGEKLLDNVYSSATKEADMGNGK
 IKYTPDNNEAEYIHYNTVTKVTNNKPDNNAKGEGITGTSISNLTGKDPYSKE
 NITVKNQGLITTKVTKETGILLTAPEPIVINSKSGDSTVNGNSITYNKKTGELVS
 PGEIIDNKGTIINGHDLVNNISGLKVEGPIPFENKADKMSGIAKEIIIKKGDDYD
 LVGPIDKAKNTTFANARYLYKDLGVHVTVPKFNDPVSSVMGVSSTATYNPQDSI
 LRGTNFMPEPDSAKQNIVLNKKRRLVGNAYLSSGKONIISQPKIVYLDTKD
 AETPTNSVINDOYTIKSTAKVNRSEGAFAKADVKSVGDNFEFSANEAKGNTSVV
 HFTGNVSKSQKQEGDVFPTGDKADLYMSKNDKYQAKVIVDTKSTFTQLNKRKIDNS
 YLELDLIKRYARKNPVLITDGPKGSTLVKADDTVGIIDKELIKLNKNVYVKNINE
 KKEETVLTAADRGVATKEMADVDYDKVKIVTKESVTTANEGHYDMVNRKIRAKGNVHDY
 TGDKTSATFNDMTATKKK"
 2447..3172
 /gene="FN0695"
 2447..3172
 /gene="FN0695"
 /codon_start=1
 /transl_table=11
 /product="ABC transporter ATP-binding protein"
 /protein_id="AAL94891.1"
 /db_xref="GI:19714222"
 /translation="MISLSADSLVKTYKKRVKVDKSVLKNKGEIVGLGPNAGKTT
 TFYMITGIVKPDGQVNCADODITNPMYKRAMDGIGYLAQFPSPVFNLTVEBIEVW
 LEMKNI SKMQRETVDRLLEFELTHVRDSILGYSLSGGERRIEIARTIANPFSILL
 DEPPAGVDPIAVEDIQNIIRHLKKRDLGILITDHNVRTLSITDKSVYIMAKGVILIEG
 TPREIANNPEARLYLGEKEFLD"
 3255..3407
 /gene="FN0696"
 3255..3407
 /gene="FN0696"
 /codon_start=1
 /transl_table=11
 /product="unknown"
 /protein_id="AAL94892.1"
 /db_xref="GI:19714223"
 /translation="MFAESSNPYKAEENVYFLMGDFVNEWGFTNFNPIDEKWKBEES
 SLOQFL"
 3553..6156
 /gene="FN0697"
 3553..6156
 /gene="FN0697"
 /EC_number="6.1.1.7"
 /codon_start=1
 /transl_table=11
 /product="Alanine-tRNA synthetase"
 /protein_id="AAL94893.1"
 /translation="MANKETIENVKIIQKSYDETPYKSKTFYFTYQPGROQMVLKJLGF
 KTPDLKARVLEIGCSFGGNIIPFALENPKAEVVGIDLSNVQIIEGNRIIEFLNLENI
 /db_xref="GI:19714224"
 /translation="MLTGNIEIREKFIETFMQKHQKHFSASLIPDDPTLLLTVAGMVP
 FKPYFLGQKEAPCPVTTYQKCIRTNDLENVGRFARHFTFFEMLGNFSGNYPFKKEAI
 KSWSEVTEVLKINKDLAVTVFTTDDAEAKIWEBCNPKERI VMGSENNWSAGP
 TSCGSPCSIHVLDLQYCGDENS KIGDEGTNRFEIWNLVFTEKRMEDGSLRPLP
 KNRIDGAGLERIAAVQSKTNNFETDLLPFIIEBAGKITGSGYKGNPETFSLKUIT
 DHARVTVLNDGVIPFSGRGYILRRIRAVRHGRGLIGYKDLFMVKMDVKVKEKPE
 IAYPDLRKNVENIRKIKYIBEEKFSNTLDQGLVNOEIDNLTNGKNKLDGDSIEFPY
 DYGVTFYGTGTEKTEDEGKLLSREAKDRKYLII FOKTPFYAESGGQVGGQGIYSQN
 RAGKVLVDOKQDIFHTVTKLEKGMPEENKTYKLEVVDVKRLDTAKNHTATHLHAKL
 REIVGTHVQOASLVNDSKLRPDPFSHEALTEEQLSKIEDIYNKKIIREGIEVVVSHHS
 IEAKGLGAMULFGDKYGDVVRVDVHGFSFTELCGGTHIDNIKIGLFIKGTISGGTAA
 GVRRIEAKTGYGAYLVKEBEADIKNIEQKLKATNSNLVKEKNELETLEKDTKELEBI
 LKQKLAIFETKAAISGMBEIGGVKYLIAAFKOKSTEDLRMTIDTIKDNNEKALIVLAS
 TODKLAFAVGVTKTLDKIKAGDLVKKLAETITGGKGGRPDFAQAGKDEGLDAFK
 EVREIIEAKLV"
 6172..6588
 /gene="FN0698"
 6172..6588
 /gene="FN0698"
 /codon_start=1
 /transl_table=11
 /product="DNA integration/recombination/inversion protein"
 /protein_id="AAL94894.1"
 /db_xref="GI:19714225"
 /translation="MKRYIALDIDGVIRIGVARSDDIMGIIASPLETINRKKVSKVRIA
 EICKENLTNLVVGIPKSLDGEKROAEKRYIEKLEKKEIENLEIIEVDERFSTVIA
 DNILKELNKAIEKRVVDKVAASIILOTYLDMMK"
 6612..7847
 /gene="FN0699"
 6612..7847
 /gene="FN0699"
 /codon_start=1
 /transl_table=11
 /product="Protein translocase subunit secD"
 /protein_id="AAL94895.1"
 /db_xref="GI:19714226"
 /translation="MNKLPFLRLIIVIAIFAVALYVSLAKPIKGLDLKGGAVVLEA
 VEDNSNVKIDNDAMNRLIEVLNRRVNGISVAGVETIQKAGDNRIIVELPGLQNTTEAI
 NLIGTKALMEFKLMNEDGSLGETLLTGSALQKAEVSDNLGRQISFNMTPEGAQUPA
 KITRENIGRQLAITLGVVQTPAKINTEISGGSGVITGNTVVEEAKTALLNAGALP
 IKAEIAETRTVGATLGDESIQSKNAGMVAI VLIWFMIIIFYRLPGIIADLAI IIFGF
 ITPACLNFTIDATLTFGIAGFILSLGMADVANIIFPERIKEELRFGNSIRNSIDSGFN
 KGFIAIFDSNLATLIITTLIFVFGTGPFGKFAVTLALGTLASMFTAITVTKVLLETFV
 NMGFRSPKLGFTVEEAK"
 7847..8800
 /gene="FN0700"
 7847..8800
 /gene="FN0700"
 /codon_start=1
 /transl_table=11
 /product="Protein translocase subunit secE"
 /protein_id="AAL94896.1"
 /db_xref="GI:19714227"
 /translation="MKTNLHVKIKIYLSISLVLTLSIVIFFTKGLNKGIDFSGGN
 LPOLKNGTTLVTLQINENLDKALKEIPQINSNRKVOISDDGTIVRVPETSENGSAG
 KVLNNLKEIGSVTLNDEKDGKASIGDLSKSAIYSGIGAILIVITWRFESFAIG
 GLSLHDLIIIAVGFTALMGYEDTFPIAAILITILGYSINDTIVIVDRIRENLKRRKH
 MWILEQCMDESINQTAIRSLNTSVTLFVSVAIILVFGASLTKTFIMTLLIGILAGTVS
 SIFVATPPVYLLNKKRKNMEDFMFKDDEENNDRKRVKILV"
 8822..10375
 /gene="FN0701"
 8822..10375
 /gene="FN0701"
 /EC_number="2.1.1.-"
 /codon_start=1
 /transl_table=11
 /product="Methyltransferase"
 /protein_id="AAL94897.1"
 /db_xref="GI:19714228"

QY 215 TCTTTGTTCCCGAGCTTCTCAGTGAAGAGACATCTCTCTGATCTTGGGAAAAACA 274
 DB 8519 AACGCCCCATTCAGGTTTAAAGAAAAAGATGAGGTCCGAGTTTGGGAAAGACAGC 8578
 QY 275 TGTACACAAAGGATATCTCAGCATGCTGTGGTGGCAATGCTTAATGTTGATGAGC 334
 DB 8579 ATTAAACTTCATTTATCGCTCTGATGGAGCTTTAATTTAGATGGGCTTTATGGCA 8638
 QY 335 GTATATATATAGATTGGAGCGCTCATCGCTTGGAGAGCTGTTCTTGAACTTTTGCTT 394
 DB 8639 CTATTATCTCATGCGGCGGATGCTGCAATGGCGTATGAGTCAATCTTTTGTG 8698
 QY 395 ATCTGGCAGCTCAGCTTCTTGGATGGCGACCTACCTCTTGAGAGCTCGCGGAGT 454
 DB 8699 ATCTGGCGGCTCATGCGATTTTGGAGCGAGCTGATTTTACGGGAGATGGCAGGAT 8758
 QY 455 GTTCTGCTATGGGAGATGGCGGTAGATGCAATGTTCTTGTATTCGAAAGATCCGAGAG 514
 DB 8759 GTTTTAAACCGTGGGATGGCGGTGATGCTATATCATATCAACGAAAGCATTAAGAA 8818
 QY 515 GAATTTTATGTCTCAAAATCTTAAAAATCTGAGAAAAAGATATCAAGGCTTTT 574
 DB 8819 GTTTTAAAGAGAGAGAGGAGGAGTGTGTAAGGCGATTCATTTAGGCTATATCAATGCGAGC 8878
 QY 575 GGAGCCATTTTGTGATCTTACCTGACTACATATTTGGCTGAGACCTCTTTCTTCTTA 634
 DB 8879 CGGCGATTTTGTGACTCCAAATATCACTCTTGATCGCTTGGGTGCTAAAGCTTAT 8938
 QY 635 GATACAGGGCCATTTAAAGAGTGTGCTTGGACATTTGATTTAGAAATTTTCTTCAATG 694
 DB 8939 GGACAGAGAGGATTTAAAGAGCTTTGGCTCCTCACACAGGACATTTGGAGTTTGGCCTATC 8998
 QY 695 TTTAACGCTCTTTCATGATTAATTTTCTTCACTGCTGATGATTAAGACCAACAT 754
 DB 8999 ATCAACCGCTATTAATAGGACGCAAG-----GGAATTTTCAAGCCCTTTTA 9043
 QY 755 ACAAGTTGCATATGATGAATAAGTCTGGGAGTAAAGCATGATTTCTTGAAGAGATGC 814
 DB 9044 CCTAATATGAGCCCAACGAAAA-----CCCTTTACTTTTGGTTGGCGTGAATAATA 9095
 QY 815 AAAAACTTTGGGCTGTTTGTGAAGTGTTTTCTTTAGGCTGCTGCTCGGGGTTT 874
 DB 9096 AAAGAGCTTAGAGAGTTTATGAAA-----TTATTTCAACAAGTTAGAACTTTAAGTTT 9149
 QY 875 GGAGCCTGGAATCCGTTTGGGAAATGATTTTAAAGAGGATGCTTTACTTTAT 934
 DB 9150 CATCGTATATTCATATTAATGAGGTGATGTTTCAAGCGATCTTGGGCTTCTGGCGTT--- 9206
 QY 935 CCAAAAGGATGGCATCAAGATGTTCTCAAAATGCTGGCAAAAGTTGTCATTAACATA 994
 DB 9307 -----AGGGCTTTTGTTTTAAAGGGTTCCTTTAGGATTTGATTTTGGCGGGGAGT 9261
 QY 995 CAGGAAGCTGCTCTTCTTTAGAGACTTCGTAATCAAACTTTGATCTTTCAGAAAAAG 1054
 DB 9262 TTGGTACAGGTGCTTATACATCAAAACCCCATTAAGAAAGTCCGCGATCTGTTGAA 9321
 QY 1055 ATCAAAATCTATTTTATGATTAAGCTTTAAGTATTAAGCAAGATGAGCTCTCTC 1114
 DB 9322 AAAGAGCTCGTTTAAAGGGGTGCAAGTGAAGCAATTTGATCTTAAAGAAATTTTA 9381
 QY 1115 CTAAATTTAAGATCATGAGCTGGCGTATTTGTGGAGTTGTTGTCAGAAACAGGCTTGA 1174
 DB 9382 ATCAAAATTCCTTTTGTGAAAAAGCCGCAAAACGAAATTTGAAGCTATGCTGCTAAC 9441
 QY 1175 TTTCTCTCGAAACTTAACGAAACGCAAAATTTTGTCAAAAGTAAAGCAGCAACTA 1234
 DB 9442 ATTCTTAAACCAGCGGCGATTTTGAATCCGTAAGTTTGACATCTTTGGCCCTAGAGTG 9501
 QY 1235 TCGAAGAAATGCGTTATCAGCGCAATCGGCGTTTAAAGAGCTTTGGCAATCATCTTG 1294
 DB 9502 GGGAGCGAATTTGAAGAAAGAGGCGATTTTGTGCTGATTTTAAAGCTTATATGAGCATG 9561
 QY 1295 CTCTATGTGAGTTGGCGCTTGAATGCAATATGCTTTCAAGTCCGTAATGGCGTTTAA 1354

DB 9562 GTTATGTAGATTCCGCTATGATAGCCGTTTCCGCTTTAGCCAGCCTTGTGGCTGTG 9621.
 QY 1355 CATGACCTTTTGGCTACCTGTGCACTTGTGTTATAGCAATTTCTTTTGAAGAAAT 1414
 DB 9622 CATGA-----TGTGATTTTATGCGCAAGCTCGGTGATCTTTTAAAGATT 9666
 QY 1415 CAATATGATTTGCAAGCCATGCTGTTAATGACGTATTTGGGGTATTCATTAACAT 1474
 DB 9667 GATATGAATTTGAAAGTATGGCGCTTGTCTACATTAATGATGGTATTCATTAATGAT 9726
 QY 1475 ACTTGTATCATTTTGTATGATGATTCGTGAAGATCGCCAAAGCAACCTGTTTACCCTTANG 1534
 DB 9727 ACGATCATTTATCTTTGATGCAATCAGAGAAATGCTCTCTCAAAAAACAAAAGCCG 9786
 QY 1535 CATGTTTATGATTAATGATGCTTCAAAAGAGCTTCAAGCCGCGCATTAATGACACAGCT 1594
 DB 9787 ACTCAAGCCATGATGAAGCATTTTCAGACCGCTCACGCGCACGCTTTTAACTTCTTA 9846
 QY 1595 ACAACTATACATTTTGTATATGCTTTGTTATAGCGCGCTCTCTGCTTTAATTTT 1654
 DB 9847 ACCGTGTTTGTGATGATTTGTGCGTGTGGAGATGAATCATTAATTTGCTTT 9906
 QY 1655 GCATTTATTTATGACCAATWAGGATTTCTTAAAGAACTTTATCGCTCTTATATTTGACCA 1714
 DB 9907 TCATTTGCCATGCTATATAGACAGATTTGATGAGACTTATAGCTCTATTTTCAATCGCCCT 9966
 QY 1715 CTTCTGTTGTTTATGATGCTGTTAA 1740
 DB 9967 AAAGTAGCTTATGTTAGGCTTTGA 9992

RESULT 8
 CUI1168X4/C
 LOCUS CUI1168X4 282183 bp DNA linear BCT 09-JUL-2002

DEFINITION Campylobacter jejuni subsp. jejuni NCTC 11168 complete genome;
 segment 4/6.

ACCESSION AL139077 AL111168
 VERSION AL139077.2 GI:6968444

KEYWORDS Campylobacter jejuni subsp. jejuni NCTC 11168.
 Campylobacter jejuni subsp. jejuni NCTC 11168.

SOURCE Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;
 Campylobacter.

ORGANISM
 REFERENCE 1
 AUTHORS Parkhill,J., Wren,B.W., Mungall,K., Kelsey,J.M., Churcher,C.,

Basham,D., Chillingworth,T., Davies,R.M., Feltham,T., Holtroyd,S.,
 Jagsels,K., Karlyshev,A., Moule,S., Pallen,M.J., Penn,C.W.,

Quail,M., Rajandream,M.A., Rutherford,K.M., VanVleet,A.,
 Whitehead,S. and Barrall,J.B.G.

TITLE The genome sequence of the food-borne pathogen Campylobacter jejuni
 reveals hypervariable sequences

JOURNAL Nature 403 (6770), 665-668 (2000)

MEDLINE 20150912
 PUBMED 10688204

REFERENCE 2 (bases 1 to 282183)
 AUTHORS Parkhill,J.

TITLE Direct Submission
 JOURNAL Submitted (09-FEB-2000) Submitted on behalf of the Campylobacter

sequencing team, Sanger Centre, Wellcome Trust Genome Campus,
 Hinxton, Cambridge CB10 1SA E-mail: parkhill@sanger.ac.uk

COMMENT Notes:
 Details of C. jejuni sequencing at the Sanger Centre are available

on the World Wide Web.
 (URL, http://www.sanger.ac.uk/Projects/C_jejuni/).

FEATURES
 source Location/Qualifiers
 1..282183

/organism="Campylobacter jejuni subsp. jejuni NCTC 11168"
 /strain="NCTC 11168"

/db_xref="taxon:192222"
 complement(43..1443)

gene /gene="Cj1009c"
 complement(43..1443)

CDS

```

/gene="Cj1009c"
/notes="Cj1009c, unknown, len: 466 aa; 29.1% identity to
HP0282"
/codon_start=1
/transl_table=11
/product="hypothetical protein Cj1009c"
/protein_id="CAB73265.1"
/db_xref="GI:6968445"
/db_xref="SPTREMBL:Q9PNT1"
/translacion="MNNILIIIDGILAKHFLERLCPEKGLGYFTVVQCNSEKNLNLI
SSEYIDHYDPSTARLENIMSKDFQAFYMODEFETKSKYEALRSNPNLEIEM
DFWGLSVNDHANLADARMTLSRFMDFLPDIALTAOYIGLVGEINVEVKIPAGSIFA
YRHLSSQQRWRVILYRNSKIYFVPSFVLENDLILVDFVLQSFPHIRGKA
QDPMPGSGNSVFADIDMKNQNNQERVLDITLKLQTKSNKARFFIHIIVPKLGWYFE
KLKLSDEKGVDFDPTDFKQISMWLQNNIDGLVTDINKFEKQAFDILKIPIM
KVGASDELEKRAIILSADSELENANAVITDLSKQDFGVILYYNPNQNTWDEE
YFRSLXLYDKNIQIINKNDENPLNLQYREDLQFVFSQKELLNRDFARNLSTNLR
HYTKRNYOLFVPE"
/length=.2594
/gene="tgt"
/length=.2594
/gene="tgt"
/EC number="2.4.2.29"
/notes="Cj1010, tgt, probable queuine
tRNA-ribosyltransferase, len: 373 aa; similar to many e.g.
TGT ECOLI queuine tRNA-ribosyltransferase (EC 2.4.2.29)
(375 aa), fasta scores; opt: 1016 z-score: 1190.9 E(): 0,
41.3% identity in 373 aa overlap. 56.7% identity to
HP0281"
/codon_start=1
/transl_table=11
/product="queuine tRNA-ribosyltransferase"
/protein_id="CAB73266.1"
/db_xref="GI:6968446"
/db_xref="SWISS-PROT:Q9PNT0"
/translacion="MEFKLHKQGMARVCVITTAHSTFLTPVPMVGTGAVKSLDAN
DMKNLDAKILANTYIMYLRKPSKVVDKFGGLHGFYKDFORSFLTDGGQAFSLSN
SKHNEGIEFASHYDGSRLHFTPKSVLDITQDFNSDITMMLDLVALPATKRVKISV
DRILWAKAEITTHKNQNGKIGIQNIFGI IQGDTYEBERKRCALSNLMPFDGLAI
GGLSVGENALMYETQNLNPDENRPVLCMGVGTPELDVENVERGVDMFDCVMPTR
NARGTPTFSFGFNIKKAFINDHVIDSTCSYCTCRNSRGVNLHFLPAKELTPFR
LASLHNLHYLELAKRRAILNNSFTQFRNFVHLRGK"
/length=.3363
/gene="Cj1011"
/length=.3363
/gene="Cj1011"
/notes="Cj1011, probable membrane protein, len: 255 aa; no
Hp match"
/codon_start=1
/transl_table=11
/product="putative membrane protein"
/protein_id="CAB73267.1"
/db_xref="GI:6968447"
/db_xref="SPTREMBL:Q9PNS9"
/translacion="MLSLHNLKILQAKNLETYSIYQKIFVYVGLNLEWLFND
EKYKLDQDGSEIEEFNSFIKLEDFKQVNTHFQNLHKEGILKGGFKNFY
KGSFVLQKINKNLQNLSEAFNLILLSEAOYKHLKILNLSILSKNTKEHLAR
IDTLYTNAIKNEKNKSIYLLSILSSIPLPLNLIVFGFMNTNLLFFKXDSPVGTLY
IFSLICLILVGFIFYYSKTKTEFDLDEGKAKQTK"
/length=.3416
/gene="trna-Arg"
/length=.3416
/gene="trna-Arg"
/product="trna-Arg"
/notes="trna Arg anticodon CGG, Cove score 83.29"
/complement(3495. .3944)
/gene="Cj1012c"
/complement(3495. .3944)
/gene="Cj1012c"
/notes="Cj1012c, possible membrane protein, len: 149 aa; no
Hp match"
/codon_start=1
/transl_table=11

```

```

/product="putative membrane protein"
/protein_id="CAB73268.1"
/db_xref="GI:6968448"
/db_xref="SPTREMBL:Q9PNS8"
/translacion="MVIFCAALALLPLGVIAVLITSDGKTTIKKOKTSOKOVAEKT
KFFDTLDKMLIAASDVKLTDIELKELAKLYVOTHLGSKTSKELDEAAKKLEFVSA
LAANINASQTVSYLNKELKRYISYKKEIDAYEHMGLAKRKIKEDK"
/complement(3961. .7206)
/gene="Cj1013c"
/complement(3961. .7206)
/gene="Cj1013c"
/notes="Cj1013c, probable membrane protein, len: 1081 aa;
contains three domains; aa 1-90 contains three membrane
spanning domains; aa 90-780 is non-membrane, and aa
780-1081 contains ten possible membrane spanning domains.
Some similarity in C-terminal membrane domain to eukaryotic
proteins e.g. CCSA, CHLRE cytochrome C biogenesis protein
CCSA (353 aa), fasta scores; opt: 459 z-score: 515.4 E():
2.1e-21, 36.0% identity in 267 aa overlap, and to
prokaryotic equivalents e.g. CCMF, ECOLI cytochrome C-type
biogenesis protein CCMF (647 aa); BLASTP scores; E = 0.20,
24% identity in 293 aa overlap. 39.0% identity to HP0378
(domains 2+3 only)"
/codon_start=1
/transl_table=11
/product="putative membrane protein"
/protein_id="CAB73269.1"
/db_xref="GI:6968449"
/db_xref="SPTREMBL:Q9PNS7"
/translacion="MNNIISIGDLRVSVVFLFALFALATFIESAYGTPTWAMV
YDTWFYEQLLGILNLCGMFRYKMGFLKPLMFIHPSFLFGLFSAMTRVAGFEG
ILPIREHTQNSLIESKTSLRISAIGDGRYSANDRYIGNLPFANSFKLNLGDDQ
AVLKQDLILNAHYTKENNNSDPLVLSQKSGQVDVKFEKVEYIEGYNFAM
NDNVKAPFVKIDENLTLSSSENHLFSLMDLQNLDLKIGKANAKERRLYEINDISFV
VKAASLHAQALSGSNRPODESFLMFKSAWLSVGRMTLJSTGEQPMKNSLLJHK
DFALSNEKLELTGSNALKLSYKNESKEFTYFKNKEIMELADQKFFISWALSY
QPPFDIYDRDFVLDYRPGSMSPASASEITVKNNENFDYRIFMNVLDYDGYRFYQ
SYVDQEKGLTVSKNDPKPTIYIGVFLCLGMFNFLNPHSRFTLARLINDTLK
HTSVITFILLSGSEKTAQDNLSTLPVNTNHAALATLIVQKSDAGRMVPPDTLS
REILEKHOSDSYKQNSNAVMSLVMDVKQLEPFLMPQNAVRDAIANILEIPS
AKIAYKDFDENRYKLOKYVENAKRKNPARGVDFKEILKUDERANVNLVPSGL
FKFIPQNNVNLAPFSAVTTKDGDEHIVALIQNYFSAVENAFKQGNWIRADSG
LKFIKEYQKIGYKVMPSKTEMEIFSNAKEIFVKLAPVYLIAGLLLILVFSKMV
PNLKSIFKVVYLVNLAPVITVIGLGRAYLSHAPNSNGYESMYVIAWLSLGI
FFRSKSPIALSLSILSGVLMVAHLSMMNPQITNLVPSVLYSLSHVSVITASYGF
LGIICALLGIPTFLMCLFKDKGVNLNLRNITEATREINEMAMTIGLCLITVGNFLGA
IWANESGRYMSHWSKETHALVSLIYAAIHLHRMPKYNQFVFLMSFPAYVVIIM
TYFGVNYFLGLSHSYAAGEAAQIPNYVYMGFALMVVLAFARRKRNFGKL"
/complement(7224. .7919)
/gene="livF"
/complement(7224. .7919)
/gene="livF"
/notes="Cj1014c, livF, probable branched-chain amino-acid
ABC transport system ATP-binding protein, len: 231 aa;
similar to e.g. LIVF ECOLI high-affinity branched-chain
amino acid transport ATP-binding protein (237 aa), fasta
scores; opt: 686 z-score: 797.6 E(): 0, 46.4% identity in
233 aa overlap. No Hp ortholog. Contains PS00017
ATP/GTP-binding site motif A (p-loop), PS00211 ABC
transporters family signature, and Pfam match to entry
PF00005 ABC tran"
/codon_start=1
/transl_table=11
/product="branched-chain amino-acid ABC transport system
ATP-binding protein"
/protein_id="CAB73270.1"
/db_xref="GI:6968450"
/db_xref="SPTREMBL:Q9PNS6"
/translacion="MLIVKDLHVYVGLIEAVKIGDPKVTGHVLSLIGSAGKSTSL
NALLNSVKRTGEVNFGLYDTRKHLTHTLVOKGIALVPEGRVFINLSVEENLKICAFN
NAENYBHLREQMYKLPFRASKKLAGLTSAGGAQMLAISRALMSFKLLMLDPEPSL
GLAPKIVGEVDFIIVRLKEEGITILLVEQNAYSALKISDYAVYLENGRIVMENAKNL
IGDDEIRKKYVLGL"

```

```

misc_feature      complement (7299..7841)
                  /gene="livF"
                  /note="Pfam match to entry PF00005 ABC transp. ABC
                  transporters, score 147.40, E-value 2.5e-40"
misc_feature      complement (7482..7526)
                  /gene="livF"
                  /note="PF00211 ABC transporters family signature"
misc_feature      complement (7797..7820)
                  /gene="livF"
                  /note="PE00017 ATP/GTP-binding site motif A (P-loop)"
gene              complement (7906..8676)
                  /gene="livG"
                  /note="Cj1015c, livG, probable branched-chain amino-acid
                  CDS
                  /note="Cj1015c, livG, probable branched-chain amino-acid

Query Match      8.0%; Score 148.2; DB 1; Length 282183;
Best Local Similarity 46.1%; Pred. No. 8.2e-23;
Matches 722; Conservative 0; Mismatches 778; Indels 67; Gaps 4;

Qy 161 TTTACCAACCGTGAAGTGAACAACTGCGCTCAATTTAAATCTGAGCGATCTTT 220
    |||||
Db 86752 TTTACTCAAGAAAGACGCGATGCGGCTAGCTTTAAGAGTGGAGCTTTCGCG 86693
    |||||

Qy 221 GTTCCGAGGTTCTCAGTGAAGACGATCTTCTGATCTTGGGAAAAACAATGACA 280
    |||||
Db 86692 CCTGTAAATTTACTTGAACAAAGACATGACCATCTTAAAGCTGATATTTAA 86633
    |||||

Qy 281 CAAGGCAATATCTCAGCATGCTGCGCTTGGCAATGCTTATGTTTGAAGCGATAT 340
    |||||
Db 86632 ATGAGTATGATAGCTTAACTGGTGATCAGTTTATATGATATTTATGATATGAT 86573
    |||||

Qy 341 TATAGATTGGAGCGGTATGCTTGGGAGCTGTTCTTGAATCTTTGCTATCTGG 400
    |||||
Db 86572 TACGTTGGCAGAAATTTTGAACAATTTGCAATGCTAGTTAAATGTTTGATGAT 86513
    |||||

Qy 401 GCACCTCTACAGTATTTGATAGCGGCACTCACTTGTAGAGACGCGGATGTTCTT 460
    |||||
Db 86512 GCCGTATAGCTATGTTTGAAGCACTTAACTTAAACAGGATGCGACCTGTTTAA 86453
    |||||

Qy 461 GCTATGGGATGCGCGTATGCAAAATGTTCTTGTATCGAAGATCCGAGAGATTT 520
    |||||
Db 86452 ACTGTGGTATGCGAGTATGCAAAATGATATCATGAGCGATATCGAAGACTCTTG 86393
    |||||

Qy 521 TTATGTTCTCAAAGCTTTAAAAATCTGTAGAAAAGATATACCAGGCTTTTGGAGCC 580
    |||||
Db 86392 CGAGATGAGTAAATATATAGTTAGTATAGAACAGGTTTAAAAAATGCTATGATGCG 86333
    |||||

Qy 581 ATTTTGAATTTCTAATCTGAATACAGTATGCGCTCAGCACTTTTCTTCTATGATCA 640
    |||||
Db 86332 ATTTATGATTTCAAAATATCACTTCTTATGTTCTGATGCGCTTATGCTTATGGAACA 86273
    |||||

Qy 641 GGGCTATTTAAAGGTTTGTCTTGTGACATTTTATAGAAATTTTCTCTCAATGTTTACG 700
    |||||
Db 86272 GGAGCAGTTAAAGGTTTGTCTGTGACCTTTGGAAATGCTATCGTTGTTCTATGATTACA 86213
    |||||

Qy 701 GCTCTTTTCATGACTAAATTTTCTTCATGCTGTGATGAATTAAGACCAACATACACAG 760
    |||||
Db 86212 GCTATTTTAGAACAATGATGATTTTATTTATTTATTAACACGATATGAAAAAATAT 86153
    |||||

Qy 761 TTGCATATGATGATTAAGTGTCTGGGATTAAGCATATTTCTTGAGAGATGCAAAAAA 820
    |||||
Db 86152 AATACAAATTTTGTGTTTATAGAGAAAGATATATACAGTTTATAGCAAAAAA 86093
    |||||

Qy 821 CTTTGGGCTGTTTTCGAAGTGTTTTCTTTTAAAGTGTCTGCTGCGGTTTGGAGCC 880
    |||||
Db 86092 TTTATGATTTTATGAAATGCTTTTGTCTGATTTCCCTTTT----- 86050
    |||||

Qy 881 TGAATTCGTTTGGGAATGATTTTAAAGAGGATATGCTTTACCTTTAATCCAAA 940
    |||||
Db 86049 -----TTTATTTTGTGTTTGTGTTATTTATTTACTTTGG 86013
    |||||

Qy 941 GAGCATGCAATCAGCATGCTTCTCAAAATGCTGCGAAAGTTGTGATTAACACTACAGAA 1000
    |||||

```

```

Db 86012 GATAGGGGTTTGCAATATGCAATGATTTTATAGCGGAGGAACCTTACTTACGCTTAATAT 85953
    |||||
Qy 1001 GCTGTCTTTCTTCTAGACATTCGCTATTTCAACATTTGGATCTTCAGAAAAATCAAA 1060
    |||||
Db 85952 GAAAAATGACAGCTCTTAAACGCAAAATTTTCAGAAATTTTGAANAATACAGGAACCTTTTCA 85893
    |||||

Qy 1061 ATCTATTTTATGATTAAGCTTTAAGCTATTAACACAGATAGAGCCCTCTCTCTAATA 1120
    |||||
Db 85892 AATTTATCTGTAACTGATGTTTGAAGCAAT--GAAAGAACTATTAACGCTTTTATAGAA 85835
    |||||

Qy 1121 TTAACGATCATGAGCTGCGCTTATGAGGATTTGTTGTCAGAAACAGGCTTATGTTCTC 1180
    |||||
Db 85834 GTATATGATATGATAGAAATGATATAGTATAGTATAGATATTAAGCACTC-----TTTAA 85785
    |||||

Qy 1181 TACGAAACTCTTAACGAAACGCAAAATTTTGTCAAAGTAAAGCAACATATGAGAG 1240
    |||||
Db 85784 AAGGATACGGGAAATTTTGAAGTGCCTGCTGATGATGATGATCCAAAGTATGAT 85725
    |||||

Qy 1241 AAAATGCTTATACAGCGACCATCGGGCTTTTGAAGCTTTGGCAATCATCTGCTCTAT 1300
    |||||
Db 85724 GAACCTTGAATTAAGGACTTATAGCTATAGCTGATCTTGTATAGCAATTTTATTTAT 85665
    |||||

Qy 1301 GTAGTTTGGCTTTGATGGAATATGCTTTCAGTCCGCTATGCGCTTATATTCATGAC 1360
    |||||
Db 85664 ATAGCGTTGGTTTATGATGCGCTTTTGTCTTGGCGCTATATATCACTGAATTCATGAT 85605
    |||||

Qy 1361 CTTTGGCTACCTGTGACGCTTGTGTTATATGACATTTCTTTTGAAGAAATTCAAATA 1420
    |||||
Db 85604 GTTGTGATTAAC-----TTTGAAGCAATTTCTTAAATTTAAATTTGATGTC 85560
    |||||

Qy 1421 GATTTGCAACCAATGCTGCTTATATGACGTTATGCGGATTCATTAACATATCTTGG 1480
    |||||
Db 85559 AATCTAGATCTTGTGCTGCTTTTATCTTACCTGCTGGAATCTTTGAATGATTAAT 85500
    |||||

Qy 1481 ATCAATTTTATGCTATTCGTGAAGATCGCAAGCAACCTGTTTACCCCTATGCAATGT 1540
    |||||
Db 85499 ATATATTTTATATGAATATCAGAAAGATATTAACAAAGTAAACCAAGCTTGACCT 85440
    |||||

Qy 1541 TTAGTATATGATGCGCTTCAAAAAGACGTTACGCCGCAAGCTTATGACACAGCTACACT 1600
    |||||
Db 85439 ATTTATATGAAAGTGTTCAGCAACTTTTATCAAGAACGTTTAACTTACAGGTTGACT 85380
    |||||

Qy 1601 CTATCACTTTTGTAAATGCTTTGTTATATAGGGGCTCCCTGCTTATTTTGAATTT 1660
    |||||
Db 85379 TTAGCAACCGTGTGATATCTTATTTCTTGGGGGAAATGATACAGATTTTCTTG 85320
    |||||

Qy 1661 ATTTAGCAATAGGATTTCTTGAAGAACTTTATGCTCTTATATATGACACACCTCTG 1720
    |||||
Db 85319 GCTTATATGATGATATTAAGGGGAACCTTAAAGCTCTATTTTGTGCAAGTCCGACA 85260
    |||||

Qy 1721 TTGTTGT 1727
    |||||
Db 85259 TTATTTAT 85253
    |||||

```

```

RESULT 9
AE000652/c 15058 bp DNA linear BCT 06-APR-1999
LOCUS Helicobacter pylori 26695 section 130 of 134 of the complete
DEFINITION Genome.
ACCESSION AE000652 AE000511
VERSION AE000652.1 GI:2314720
KEYWORDS Helicobacter pylori 26695.
SOURCE Helicobacter pylori 26695.
ORGANISM Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
          Helicobacter.
REFERENCE 1 (bases 1 to 15058)
Tomb,J.-F., White,O., Kerlavage,A.R., Clayton,R.A., Sutton,G.G.,
Fleischmann,R.D., Ketchum,K.A., Klein,H.P., Gill,S.,
Dougherty,B.A., Nelson,K., Quackenbush,J., Zhou,L., Kirkness,E.F.,
Peterson,S., Loftus,B., Richardson,D., Dodson,R., Khalak,H.G.,

```

Glodek, A., McKenney, K., Fitzgerald, L.M., Lee, N., Adams, M.D., Hickey, E.K., Berg, D.E., Gocayne, J.D., Utterback, T.R., Peterson, J.D., Kelley, J.M., Karp, P.D., Smith, H.O., Fraser, C.M. and Venter, J.C.

The complete genome sequence of the gastric pathogen *Helicobacter pylori*

Nature 388 (6642), 539-547 (1997)

97394467

9252185

2 (bases 1 to 15058)

Tomb, J.-F., White, O., Kerlavage, A.R., Clayton, R.A., Sutton, G.G., Fleischmann, R.D., Ketchum, K.A., Klenk, H.P., Gill, S., Dougherty, B.A., Nelson, K., Quackenbush, J., Zhou, L., Kirkness, E.F., Peterson, S., Loftus, B., Richardson, D., Dodson, R., Khalak, H.G., Glodek, A., McKenney, K., Fitzgerald, L.M., Lee, N., Adams, M.D., Hickey, E.K., Berg, D.E., Gocayne, J.D., Utterback, T.R., Peterson, J.D., Kelley, J.M., Cotton, M.D., Weidman, J.M., Fujii, C., Bowman, C., Matthey, L., Wallin, E., Hayes, W.S., Borodovsky, M., Karp, P.D., Smith, H.O., Fraser, C.M. and Venter, J.C.

Direct Submission

Submitted (06-AUG-1997) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA

3 (bases 1 to 15058)

Direct Submission

Submitted (17-MAR-1999) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA

Location/Qualifiers

1. 15058

/organism="Helicobacter pylori 26695"

/strain="26695"

/db_xref="taxon:85962"

complement(111..968)

/gene="HP1538"

/note="This region contains an authentic frame shift and is not the result of a sequencing artifact; similar to GB:J03176 PID:152084 SP:P51131 percent identity: 29.8%; identified by sequence similarity; putative"

complement(965..2203)

/gene="HP1539"

complement(965..2203)

/gene="HP1539"

identity: 39.09; identified by sequence similarity; putative"

/codon_start=1

/transl_table=11

/product="ubiquinol cytochrome c oxidoreductase, cytochrome b subunit (fbch)"

/protein_id="AAD08579.1"

/db_xref="GI:2314721"

/translation="MASIKKAKNLGELMDRLGNLKVLMTEYIPKINFLWANG VILLTGLVGVSGIFLLMYKPKAKAFDSVNETIQEYVAYGHLWPHMHTAASMF VLIYHMFVGIYSGYKGRMISCHILFVFSAAFGSYMLPWQMSYAAAVIT NLFQGEIFIGADVVEVIRGNVADSTLTFPLHVLFLPIAIIILLVGVHYSIRIPH VNNQGEIIFDEBEKFIQEKKEKSVIPFPWFVLSKDIFFVCAFWFFFYLVYHY DFANDPINFERNLSKTPPHIYPEWFLWSYEVIRGFFFSADLGLMAFGVAQVIFFL PFLDRSPVAPAKRPAFPMVFWLVIDMIVLTIYVGLPKIGIKYGLAGSITFFAL FFVLPITTAESKKQGV"

complement(2214..2717)

/gene="HP1540"

complement(2214..2717)

/gene="HP1540"

/note="similar to GB:J03176 PID:152083 SP:P51130 percent identity: 39.22; identified by sequence similarity; putative"

/codon_start=1

/transl_table=11

/product="ubiquinol cytochrome c oxidoreductase, Rieske 2Fe-2S subunit (fbcf)"

/protein_id="AAD08580.1"

/db_xref="GI:2314722"

/translation="MADIQRDRFLGMSLAVTAIGAIASLVAMKKTWDPFLPSVVSAGF

TTTIDVNMORGGQSTVWROKPVYILKRSKKEGFNEKRDPKVGESVPTTATQICTHUG CITYQDDEEKGFLCPCHGSRFTSDGVNIAGTTPPPRPDPPIPPKIEGTKITTFGEAGAY KQWMAKA"

gene

CDS

complement(2842..5841)

/gene="HP1541"

complement(2842..5841)

/gene="HP1541"

/note="similar to GB:L42023 SP:P45128 PID:1007188 PID:1221387 PID:1205499 percent identity: 37.66; identified by sequence similarity; putative"

/codon_start=1

/transl_table=11

/product="transcription-repair coupling factor (trcF)"

/protein_id="AAD08581.1"

/db_xref="GI:2314723"

/translation="MIQSSYRALNKGFDYOILACKDFKESELAKEVISYFKPNTKAI LFPEFRACKNDLRSFPEEFLOLLGGLREFVQALENKQETIIITAPIALLHPLPKKEL LESFKITLLKYNLKKOLKDLFYGYGIELDLVEVEGASFRGDIVDIYAPNSKAYRLS FDTCEESIKFDPITOMSLKDELLEIPIPTLPSLDESSYKDLTKVEQSPUNFSK DLTSFGLWFLGKAKOILLIYKSIISPRALEEIQELASNELDCERCFKFIAPFVLFNFK YEDLEIHAHALEGFIALHSHNKITLLAPNKITILDNAISALDAGNMECVIAPFVLFNFK PDGIFISLNSFERKKKROKSKALNELNPGCEWVVDYGVGVFSOLVQHSVLGSKRDF LEIAYLGEDKLLLPVENLHIIARYVAQSDSVPAKORLKGSKFLKAKVTKLLEIAS XIIEAARNLIIKGMVDMHLEAVFKSHAGFEYTSDDQKATAEISJSSHRVMDR LLSGDVGFGKTEVAMHAIFCAFLNGFOSALVPTTLLAQHGFETLRARFENFGVKKAR LDRYASEKNNKLLKAVELQVDALIGTHAILGAKFNGLVVDVEEHFGVQKKEALKE LSKSVFLSMSATPIPTLNMALSOIKGISLKTPTDRKPSRTFLKEKDELLEKEII YRELRRNGOIFYTHNHIASTLVKTKLEDII PKLKIALLHSHQINANESSEIMLEFAGK NYOVLICTSIVESGIHPLNANTIIIDNAQFGLADLHOLGRVGRGKGEFCFYFLIED QKSLNQALARKLALKEKNSYLGSGESVAYHDLEIRGGNLLQDQOSKKEGICNYALTY RMLDEAIYELSGKKRLEKSEIOLGVSAFNLPELIASLRLDLRYRLSCENTDEV QIHEIEDRFKGIIDLSAQFLQIITLKILANQLGIITKLSNPNQNTITYSDEKKSLE KAPSKDDNDILETLKHLRAQISLKR"

complement(5845..6255)

/gene="HP1542"

complement(5845..6255)

/gene="HP1542"

/note="hypothetical protein; identified by GeneMark; putative"

/codon_start=1

/transl_table=11

/product="H. pylori predicted coding region HP1542"

/protein_id="AAD08589.1"

/db_xref="GI:2314731"

/translation="MAIFDNNKNSAKAKTGPATIIAGTKIKGELHLDYHLHVDGELE GVHVSSTVIGOTSGVSGEIFTNKLVSCKFTGTVEAEVVEIMPLGHLDGKISSQEL VVERKGLIGETRPKNIQGALLINEQEKKIENK"

complement(6174..7112)

/gene="HP1543"

complement(6174..7112)

/gene="HP1543"

/note="similar to GB:U07173 PID:460955 PID:1100877 percent identity: 37.15; identified by sequence similarity; putative"

/codon_start=1

/transl_table=11

/product="toxR-activated gene (tagE)"

/protein_id="AAD08582.1"

/db_xref="GI:2314724"

/translation="MFLDRRLIYMTDSKGSRYINVHILFRQIGLYALLSVGSLFLF GISLLVNLQEKIDKQHALITKEFEKKETNEKLSLQMDFLDQLSGELNDLEE VGVNRPPEERKEGNSRRLDVAGITGLQKSFIMRLIPNDYPLESVGSAAPNKRTH PIHLVHNHTGLDLSTAINTPVVASAGVGLASKGWNGYGNLIKVFHPFGKTYVA HLNKIVYKTCGEFYKKGQLGYSGNTGMSGTGPHLHYEVRLDQPINPMSFTNNMKDPE EVFNKRSRWQSLITIINKMQDQRLSSLKAQK"

complement(7121..8047)

/gene="HP1544"

complement(7121..8047)

/gene="HP1544"

/note="similar to GB:U07173 PID:460955 PID:1100877 percent identity: 31.25; identified by sequence similarity; putative"

/codon_start=1

```

/transl_table=11
/product="toxR-activated gene (tagB)"
/protein_id="AAD08583.1"
/db_xref="GI:2314725"
/translation="MPQNLVITIDESGSKQKFSKLNKRLIISVIFLLVIGLV
GFLKPLAKMDTWAEARNALRDRPDLVQKNYLTKEIKKREELFVQKIKTISL
IEVRKANGCVHLYDEVDLDNLAAQHILMLIPNMPKITYSAIKPTERRHPKIK
IKGVESGIEFIAPLANTPVYASADGIYDFVATNSNVGKLVRIEHAQRESITTHLDH
VNVQPSFIOKGQILIGSGKSGNSGGKELHYEVRFQKLIIDACKFLAMDHPQSALE
ENKFEIKWKLFWYLEDIVQLQEHVDKALISQ"
complement(8047..9231)
/gene="HE1545"
complement(9221..9733)
/gene="HP1546"
/feature="similar to GB:L42023 SP:P43775 PID:1007194
PID:1221390 PID:1205502 percent identity: 35.17;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="folylylpolylutamate synthase (folC)"
/protein_id="AAD08584.1"
/db_xref="GI:2314726"
/translation="MKNSPLNGLNGLKAFLETPKREYHKEPSPRFIOYDFQNAFEE
IOAKVHVGTNGKSGTGRFLTLILDONKFLVHPSPHYEEREPFLNGSVVGSV
LENAHQOLSHAFSSACSVEYATILAVMLAKODVLYEAGIGEDSTNALKTIS
VTFPIIDHKEFIADSLIESIAQTKLRAMGSLIAPQOEVLVNAOKIAKEXAKLIV
VONEISKVRDYLERYHLARFLANMEVLAKEPDLPCNKOEVLNKLFNLIGICE
LPSNPILDIVGNPHSAKALKEIKRIFNKIILILYNCYGDKDAFLVLEILKFIKRY
LILEHEERVIKLEKLGILETGLFEVALFEDVEENENYLVGSPVANAIFYRYOEK
RD"
complement(9221..9733)
/gene="HP1546"
complement(9221..9733)
/gene="HP1546"
/feature="hypothetical protein; identified by Genemark;
putative"
/codon_start=1
/transl_table=11
/product="H. pylori predicted coding region HP1546"

Query Match      7.9%; Score 146.4; DB 1; Length 15058;
Best Local Similarity 53.7%; Pred. No. 2.8e-22;
Matches 303; Conservative 0; Mismatches 261; Indels 0; Gaps 0;

QY 155 GGAATTTTACCAACCGTAGAGAGCAAACTGCTCAGATTAAATCTGAGCGCATG 214
      |||||
Db 14132 GGAATTTTACCGTAGAGAGCAAACTGCTCAGATTAAATCTGAGCGCATG 14073

QY 215 TCTTTTGTCCGAGTTCACAGTGAAGACGATCTCTTGATCTTGGAAAAACAA 274
      |||||
Db 14072 AGCGCTCCCATTCAGGTTTGAAGAAAAAGATTATAGGCCCAAGTTTGAAGAAAGACAGC 14013

QY 275 TGTACACAAGGCAATTAATCTAGCATGCTGTGGCATGCTTATGTTTGTAGAGC 334
      |||||
Db 14012 GTTAAACCTTCATTAATGCTCTAGTGGGAGCTTTATTTAGATGGGCTTTATGATG 13953

QY 335 GTATATATATAGATTGAGAGCGTCATCGCTTGGGAGCTGTTCTTGAAATCTTTTGCT 394
      |||||
Db 13952 CTTTATTACTCTATGGCGGAGGTATGCTGTTGGGCTTAGTGTCATCTTTTGTG 13893

QY 395 ATCTGGGAGCTTACAGTATTTGGATGGCCACTGCTGTGACGACTGGCTGGGAT 454
      |||||
Db 13892 ATTGTGGGAGCTACAGATTTTGGAGCGACGCTGACTTTACGGGGAATGGCGGAT 13833

QY 455 GTTCTGTCTATGGGAGTGGCGTAGATGCAATGTTCTTGATTGAAAGATCCGAGG 514
      |||||
Db 13832 GTTTTAAACCGTAGGAGATTGCCGTATATCATCAACAGCGCATTTAGAGA 13773

QY 515 GAATTTTATTTGTCTCAAGTCTTAAAAATCTGTAGAAAAAGATATATCAAGGCTTT 574
      |||||
Db 13772 GTCTTAAGAGAAATGAGGAGATGCTTAAAGGATCCATTTAGGCTATATCAATGCGAGC 13713

QY 575 GGAGCCATTTTGTATTTAATCTACGATGACGATTTGGCTCAGCATCTTTCTCTCTA 634

```

```

Db 13712 CGGGCATTTTATTTATTTAATCACTTCTTGATGCTTCAAGTATATACGTTAT 13653
QY 635 GATACAGGGCGCTATTAAAGGTTTGTCTTGACATTTGATTTGAAGATTTCTTCTCAATG 694
Db 13652 GGCACAGGAGCGATTTAAAGGCTTGGCCCTTAACACAGCATTTGGATTTAGGCTCTAAT 13593
QY 695 TTACGCGCTCTTTTCAATGACTAA 718
Db 13592 ATCACCGCTATTGTTGGACAGCA 13569

```

```

RESULT 10
AP003192/c
LOCUS      301550 bp      DNA      linear      BCT 10-JUL-2002
DEFINITION Clostridium perfringens str. 13 DNA, complete genome, section 8/10.
ACCESSION AP003192
VERSION    AP003192.2 GI:18146740
KEYWORDS
SOURCE
ORGANISM  Clostridium perfringens str. 13 (strain:13) DNA.
          Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
          Clostridium.

```

```

REFERENCE
AUTHORS    Shimizu,T., Ohtani,K., Hirakawa,H., Ohshima,K., Yamashita,A.,
          Shiba,T., Ogasawara,N., Hattori,M., Kuhara,S. and Hayashi,H.
          Complete genome sequence of Clostridium perfringens, an anaerobic
          flesh-eater
JOURNAL    Proc. Natl. Acad. Sci. U.S.A. 99 (2), 996-1001 (2002)
MEDLINE    21664373
PUBMED     11792842
REFERENCE   2 (bases 1 to 301550)
AUTHORS    Shimizu,T.
TITLE      Direct Submission
JOURNAL

```

```

COMMENT
FEATURES
source
On Jan 14, 2002 this sequence version replaced gi:18145488.
Location/Qualifiers
1..301550
/organism="Clostridium perfringens str. 13"
/strain="13"
/db_xref="taxon:195102"
/note="anaerobic pathogen for gas gangrene"
complement(24..413)
/gene="CPE1825"
complement(24..413)
/gene="CPE1825"

```

```

/note="129 aa, similar to sp:YOHY BACSU HYPOTHETICAL 14.7
KDA PROTEIN IN ACCO-FOLD INNERGENIC REGION from Bacillus
subtilis (135 aa); 40.7% identity in 113 aa overlap. Also
similar to p1r:JC2527 alkaline shock protein from
Staphylococcus aureus. 2 putative transmembrane regions
were found by PSORT."
/codon_start=1
/transl_table=11
/product="probable alkaline-shock protein"
/protein_id="BAB81531.1"
/db_xref="GI:18145489"
/translation="NMENMRDANIGYINISDEVIGVAGIASRIDGILEINHNST
GIGHKFSKSLGKIKVNESEAYIEGVTQYGIKIPDVVSQVQENRIVEALIG
LKVAVLVIVQNIILIKEDKNEILIK"
complement(498..995)
/gene="SP0111AH"
complement(498..995)
/gene="SP0111AH"

```

```

/note="165 aa, similar to sp:S3AH BACSU STAGE III
SPORULATION PROTEIN AH from Bacillus subtilis (218 aa);
22.4% identity in 143 aa overlap. Putative N-terminal
signal sequence was found by PSORT
CPE1826"

```

```

/codon_start=1
/transl_table=11
/product="stage III sporulation protein AH"
/protein_id="BAB81532.1"
/db_xref="GI:18145490"
/translation="MNRKQAGIILTLALIVCTGLATRVNNOIKESMGVPTAFGEN
DDNATERSSNYFFESNRLEKQSKTIDNLKAIVEDKNTSAEQKEAEKELTEKTWAR
DYETRIELSLKSGYEDVIFIDGNKRVVVKINEELTQEKWVEIQDINVNNSKVYDV
DIEKK"
gene
CDS
complement(1021..1605)
/gene="spolIIAG"
complement(1021..1605)
/gene="spolIIAG"
/notes="194 aa, similar to pir:B69712 mutants block
sporulation after engulfment spolIIAG from Bacillus
subtilis (229 aa); 28% identity in 157 aa overlap.
Putative N-terminal signal sequence was found by PSORT
CPE1827"
/codon_start=1
/transl_table=11
/product="stage III sporulation protein AG"
/protein_id="BAB81533.1"
/db_xref="GI:18145491"
/translation="MSKILKQKITNLIILLVIMFVLVSVFTGVNNITKSEKTNL
EKYSKEMDNSQKDSVLVSQYQEKDLEKLGKNGVSDVVIINFSQSEKVPVAV
DNSQKSTTSETSEGTRVNSQETDGDKIVNSNGSBFVILKTEKPEVLGVWVVA
EGAPDSKYEITKAISSLYNISVDKYNVLMKK"
gene
CDS
complement(1641..2033)
/gene="CPE1828"
complement(1641..2033)
/gene="CPE1828"
/notes="130 aa, no significant homology"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB81534.1"
/db_xref="GI:18145492"
/translation="MNKDIDETVQEKISDVTRSDEKKSLYDEEAILKSVENKLEKS
LKDFEYNEFEVNLKIDFIDVFKFIEKAEITVLDKVKVQDVVVVGKVKVKBK
KDSFLKVEKVFKELEISHENIIVSYA"
gene
CDS
complement(2207..3379)
/gene="spolIIAE"
complement(2207..3379)
/gene="spolIIAE"
/notes="390 aa, similar to gpu:AP001516 222 mutants block
sporulation after engulfment from Bacillus halodurans (397
aa); 26.4% identity in 333 aa overlap. Putative N-terminal
signal sequence and 9 putative transmembrane regions were
found by PSORT.
CPE1829"
/codon_start=1
/transl_table=11
/product="stage III sporulation protein AE"
/protein_id="BAB81535.1"
/db_xref="GI:18145493"
/translation="MKKLIINILLSLFVFCNLKVLANDNLSSKEKIDNDSRIERF
YDINNLETEDIGNNAKEYIMNYLESGEDPITLKKIGASILSYIFRELNTVLKPI
ASLIVALLSALLKQADAFIEGVTOIAFAFCYALLIMLTSPFISINLCKEVLIT
SIIDFMNIVPVLMTATSGVTSATIDPILGAVSITPRIYTFELFLIAYFTL
QFVNLSSEKIDRMCKEFTQVNMISQFILTIFVGLTURGATADTLDAVAVTKVF
AVDNFVPSKAFSDAITVAGVSLAMKSVITSLGVILVVIYVPIIKLGLMAICPK
LTSVIEPVDSKISSVETVGEALFMINSCTISVSIMFFIMASMTAGNFIVGG"
gene
CDS
complement(3390..3776)
/gene="spolIIAD"
complement(3390..3776)
/gene="spolIIAD"
/notes="128 aa, similar to sp:S3AD BACSU STAGE III
SPORULATION PROTEIN AD from Bacillus subtilis (133 aa);
34% identity in 103 aa overlap. S.D. unclear. Putative
N-terminal signal sequence and 3 putative transmembrane
regions were found by PSORT.
CPE1830"
/codon_start=1

```

```

/transl_table=11
/product="stage III sporulation protein AD"
/protein_id="BAB81536.1"
/db_xref="GI:18145494"
/translation="MDVVIQVTSFVLISLFLVLKDNKSTPAVFLLLGAGIVILFV
MPVNEIIAPMODIARESMDSSYIAIVWKIATIAVLSFTCSYICNDAGVTLGKKVE
FYKIMILLAIPIPMANILNLSLSIM"
gene
CDS
complement(3864..4061)
/gene="spolIIAC"
complement(3864..4061)
/gene="spolIIAC"
/notes="65 aa, similar to sp:S3AC BACSU STAGE III
SPORULATION PROTEIN AC from Bacillus subtilis (68 aa);
49.2% identity in 63 aa overlap. Putative N-terminal
signal sequence and 1 putative transmembrane region were
found by PSORT
CPE1831"
/codon_start=1
/transl_table=11
/product="stage III sporulation protein AC"
/protein_id="BAB81537.1"
/db_xref="GI:18145495"
/translation="MSDLSLIQIAGVIGVILDKVLDOSGKKEATLANIVGVII
LTMIIQLISRLPSSVKSMPFLF"
gene
CDS
complement(4137..4652)
/gene="spolIIAB"
complement(4137..4652)
/gene="spolIIAB"
/notes="171 aa, similar to gpu:AP001516 225 mutants block
sporulation after engulfment from Bacillus halodurans (170
aa); 34.5% identity in 145 aa overlap. Putative N-terminal
signal sequence and 1 putative transmembrane region were
found by PSORT
CPE1832"
/codon_start=1
/transl_table=11
/product="stage III sporulation protein AB"
/protein_id="BAB81538.1"
/db_xref="GI:18145496"
/translation="MKVFFLLIIVLASSLIGLYGFGFRNLSQLRELKRALIDFEND
IVYVTPLESIESIALKAKSPIKELFNEISFKLNNEVENVYMAKESINEHKENN
LRNDFEILDLSKSGEIVNEGQIKIFNLAKELKIDLEIAEDECNKNTKVRYLGV
AVGAMIAFLV"
gene
CDS
complement(4799..5683)
/gene="spolIIAA"
complement(4799..5683)
/gene="spolIIAA"
/notes="294 aa, similar to sp:S3AA BACSU STAGE III
SPORULATION PROTEIN AA from Bacillus subtilis (307 aa);
40.3% identity in 293 aa overlap
CPE1833"
/codon_start=1
/transl_table=11
/product="stage III sporulation protein AA"
/protein_id="BAB81539.1"
/db_xref="GI:18145497"
/translation="MKEIPAVLPDKTNSCLKDKSNLKLQEIIRVKVKPLNVLDNTE
TIENVYIREDVKAIQKISNYSLVAFEDIRQGYITIOGHRVGLAGOCVIEDNSIK
TINITSINIRVCREIVGCSNLSLVENNRNNTLIISPPKCGKTLTDRITNIS
NGSIOGLKGRKTVTIDERSEAACYNGIPQNMVGRDTPVDYDNCIKSEGMMMAVRGLS
PEVICDEIGTYDKMEGLMAYNSGVSIATLHGRNVEBLYRPRVPREIVENNINKV
VLSGKKGTGTEGYNV"
gene
CDS
complement(5882..6244)
/gene="CPE1834"
complement(5882..6244)
/gene="CPE1834"
/notes="120 aa, no significant homology"
/codon_start=1

```

Query Match 7.8%; Score 144.8; DB 1; Length 301550;
 Best Local Similarity 52.3%; Pred. NO. 4.7e-22;
 Matches 320; Conservative 0; Mismatches 292; Indels 0; Gaps 0;

QY 134 AAAATCATGCGAGTGTCTCAGGAAATTTTACCAACCGTGAATGAGCAAACTGGCTCA 193
 Db 122306 AATGTGAAAGCTATATTTACAGAAATAGATCAATGAAAGAGCTGAAAGATCTGGA 122247
 QY 194 GATTTAAATCTGGAGCATGCTTTTGTCCCGAGGTTTCAGTGAAGACGATCTCT 253
 Db 122246 ATATTAATATGCTGTGCACTTACCACTTCCAGTAAAGCAAGTTCAGTTGACCTGTAGT 122187
 QY 254 TCTGATCTGGGAAAAAACAATGACACAGGCAATATCTGACATGCTGTGGCA 313
 Db 122186 GCTCAATAGGTGCAAAATGCTCTTCCAATGCTCTGAAAGCAGGCAATAGGTGTAGCT 122127
 QY 314 ATGCTATATGTTTGAAGAGGTAATATATGATTTGGAGCGCATGCGTGGAGCT 373
 Db 122126 ATAAATCTTATTTATGATACCTATCTATAGAGTTCAGGATTTATGACATGATGCA 122067
 QY 374 GTTCTTCGAAATCTTTGCTATCTGGGCACTCTACATTTTGGATGCCCCACTCACC 433
 Db 122066 CTAAGCTTTTATATATTTATGCTTTATATTTGCAATTAAGTAGAGTAACTCTTACT 122007
 QY 434 TTGTGAGACTCGCTGGGATGTCTTCTGCTATGGGATGCGCTAGATGCAAAATGCTCT 493
 Db 122006 TTACAGGATATAGCAGCTTTCTTATTAACAGTAGGATGCGAGTAGATGCTAACGTA 121947
 QY 494 GTATTCGAAAGATCCGAGAGGAATTTTATTTGCTCAAGTCTTAAATAATCTGTAGAA 553
 Db 121946 ATATTTGAAAGATCAAGAGAACTTGAATGCTAGAACTATTAACATCAGCTATGAAA 121887
 QY 554 AAAGGATATACCAAGGCTTTTGGAGCCATTTTGTATCTAATCTGACTACAGTATTTGGCC 613
 Db 121886 ATAGCTTTAGTAATGCTTTTAAGATCAATATGATTTCTAATATTAACAACCTTATGCT 121827
 QY 614 TCAGCACTCTTTCTTCTCCAGATACAGGCGCTATTAAGGGTTTGTCTTGAATGATTT 673
 Db 121826 GATTAGTTCTATACTCTTTGGATCAGGACAGGATTAAGGATTTGCTTAACTATATTA 121767
 QY 674 TTAGCAATTTCTCTCAATGTTTACGGCTCTTTTCAATGATTAATTTTCTCATGCTG 733
 Db 121766 ATAGGATATGTAATAGTATGTTTACCGCTATTAATTAATGACAAATCTTATGATTTA 121707
 QY 734 TGGATGAATAG 745
 Db 121706 GGATTCACATG 121695

RESULT 11
 AE007729/c 12592 bp DNA linear BCT 27-JUL-2001
 LOCUS Clostridium acetobutylicum ATCC824 section 217 of 356 of the
 DEFINITION complete genome.
 ACCESSION AE007729 AE01437
 VERSION AE007729.1 GI:15025281
 KEYWORDS
 SOURCE Clostridium acetobutylicum.
 ORGANISM Clostridium acetobutylicum.
 Clostridia; Clostridiales; Clostridiaceae;
 Clostridium.
 1 (bases 1 to 12592)
 Nolling, J., Breton, G., Omelchenko, M. V., Markarova, K. S., Zeng, Q.,
 Gibson, R., Lee, H. M., Dubois, J., Qiu, D., Hitti, J., Wolf, Y. I.,
 Tatusov, R. L., Sabatne, F., Doucette-Stamm, L., Soucaille, P.,
 Daly, M. J., Bennett, G. N., Koonin, E. V. and Smith, D. R.
 Genome sequence and comparative analysis of the solvent-producing
 bacterium Clostridium acetobutylicum
 J. Bacteriol. 183 (16), 4823-4838 (2001)

JOURNAL
 MEDLINE 21359325
 PUBMED 11466286
 2 (bases 1 to 12592)
 Childress, D., Zeng, Q. and Smith, D. R.
 Direct Submission.
 TITLE Submitted (24-JUL-2001) GTC Sequencing Center Production,
 Journal Finishing, and Bioinformatics teams, Genome Therapeutics Corp., 100
 Beaver Street, Weltham, MA 02453-8443, USA

FEATURES
 source
 1..12592
 /organism="Clostridium acetobutylicum"
 /strain="ATCC 824"
 /db_xref="ATCC:824"
 /db_xref="taxon:1488"
 complement (83..985)
 /gene="CAC2277"
 /note="SecP"
 complement (83..985)
 /gene="CAC2277"
 /codon_start=1
 /transl_table=11
 /product="Preprotein translocase subunit SecP"
 /protein_id="AAK80234.1"
 /db_xref="GI:15025282"
 /translacion="MLKVIKHKVWFGISLTIIPLGIAFMIYRGVTTGKLEFGIDFV
 GGSMDIKIGSFNKTVDYKIVDKYIKDSTITPTVNTVEIKSMTMDAIKMKIFDM
 OKKYNKSKALVSOKIIGTIGDELKTXALIALALVILLYIAMRFFKRGVAMI
 SIYHVDVTLAYVYGIFALPLNSPTIAGMLTIYGYKMSDITIVFDRIRESRGRHMRST
 SEIADPSIETLRSIYVTLVTLIAVSVHVFPSVRDFTPLVLGIVSGCYSSIFIA
 SPLWVIFKERRANKKLOIKAEAKK"
 complement (985..2238)
 /gene="CAC2278"
 /note="SecD"
 complement (985..2238)
 /gene="CAC2278"
 /codon_start=1
 /transl_table=11
 /product="Preprotein translocase subunits SecD"
 /protein_id="AAK80235.1"
 /db_xref="GI:15025283"
 /translacion="MKKGRSTIFPISVLIAGLAIVLSTGLANGDRIKSFPSID
 KGLDLOGVSVVEEIQKVNSDAMRTLELMLRVNKGVSSTVSRRVNNKIEIDVP
 GMYNDEVINKIGTGLKLFVDPDKTTLTSDVKKATAGTDPNTNQPILISQLNSG
 TKKFPADATOKPLGOKISIIYMDPLSDPVDVITGNAQIIGNSIIOAPQIANIN
 SGALPVLKVSOKTVGASLSALPNSTIAGAVACIIELFMILVYRIPIGMADIAL
 VLFVTVLGATFPAVAVTTLTSLIAGLLITGANVANDANVIFRRFBEELKIGTVASAF
 DAGFPHAMSIIIDSNTTITISGVILPALDSGEVKGALTLVIGVILSVTALITVRL
 LMAVIMDGLINKASHFGVNLKESGK"
 complement (2493..3857)
 /gene="CAC2279"
 complement (2493..3857)
 /gene="CAC2279"
 /codon_start=1
 /transl_table=11
 /product="Heme biosynthesis (hcrJ-2) family protein"
 /protein_id="AAK80236.1"
 /db_xref="GI:15025284"
 /translacion="MSLVKFTODDORYIIDVNSGSIHEVDENVYDLIDEDKIRSKK
 LLDRLNKYKEEISAVDEIKELIDEGVYSKDLIEDIAKSTSPSTIKALCLNTA
 HDONLCKCYCFADBEYKCGCRIMSEVAKKSIDVYIKSGPRKNIENVLFGEGPLMA
 FDIKQVVEYAKKEKYNKINIRFTWTNATLNDIMDYLDKMGNIILISFGREKLV
 NPKVRVADGTGSHDAIILKNIKRVARBDKSKOYVYRGFTLRNNDIEFDVYIADMG
 FKESIEPVLDPDNLISREEDIPIQIEQYDKLYKMDIKRKEGEIKFPHENDIN
 GGPVYKRLSGCGAGHEVYSITPSGDVYFCHQPVGNEDPKMTGIIIDDDINKMENEF
 RSHATYNAKPKRCVMAFRYCSGCGQANNYFNGDMHPIYEIGCKKQKKRICEAIALKA
 QWVD"
 complement (4317..4688)
 /gene="CAC2280"
 complement (4317..4688)
 /gene="CAC2280"
 /codon_start=1
 /transl_table=11
 /product="Predicted membrane protein"
 /protein_id="AAK80237.1"
 /db_xref="GI:15025285"
 /translacion="MEKSKTIYVSAAGVYRAIPVTLVTLILFSPVSTKISSEGITN
 MWLVTLTSLVMYSGIYSRSKSGKGLNGLVGLVYLIFVIVSVLVDSSSGMQLRD
 FIRVTLAIVGLISGLGINI"
 complement (4778..5062)
 /gene="CAC2281"
 complement (4778..5062)

gene
 CDS


```

/gene="CAC2281"
/codon_start=1
/transl_table=11
/product="Uncharacterized secreted protein, YajC family"
/protein_id="AAK80238.1"
/db_xref="GI:15025286"
/transl_table="MPNITLILVYVLLGFMFYVYVYIPRAKKKKKMKMLSEMRVND
KVVTGGIMGQITNIQOFVLIITQGPDKVIKIKKEAIAVETRAKDNQK"
/complement(5095..6225)
/gene="CAC2282"
/feature="tgt"
/complement(5095..6225)
/gene="CAC2282"
/transl_table=11
/product="Queuine tRNA-ribosyltransferase, tgt"
/protein_id="AAK80239.1"
/db_xref="GI:15025287"
/transl_table="MYTLLKTKGVRGRFETVGTETVPFMVNGTLLAAIKGAVSSM
DLKINCOVELSNTYHLSRPGDDVVKLGLKFMNWRPILTDSGGFQVFLAEIR
KIKEGYVFNHIDGRKIFMGPESMKIQSNLASTIAMFDECVENSPREYVERVSVE
RTTLMVRCKEELRLSLPDTINKKQMLFGINQGVYDDIRIEHAKTIAKMDLDGVA
IGGLAVGTHEEMVRIIDAVVPHLPQDKPIYLMGVGTSPNILEAVSRGVDFDCLVLA
RNGRHGVFTKHGKINLNEKPELDASPIDEGCQCPTCKHYSRAYIRHLFKAKENLAM
RLCVHLNDFYNTLMEDIRNAIDACDFEFPKNEKLKAWAGKA"
/complement(6243..7268)
/gene="CAC2283"
/feature="queA"
/complement(6243..7268)
/gene="CAC2283"
/transl_table=11
/product="Queuine synthetase, queA"
/protein_id="AAK80240.1"
/db_xref="GI:15025288"
/transl_table="MQLEDFYILPEELIAOHPCDKDECELMVVDKSKSEVEHRIFK
DVIDLPGDCCLVNDTRVLPARLIGNKVGTTGGKMEFLLRIDKIDKTWETLVKPKKA
KVGAFFGGVGLKAVIKSIGEDGNRVVEFYEGIFEEILDKLQGMPLPPYITBELEN
KEEYQTVYKNGKGAAPTAGLHFTNELLDRIRAKIKTVPFLHVLGIGTFRPVKVEN
IESHHMSEYMLSKESADTINERKNGGRIISVGTSTRLETIADEKGNVRQSGW
TDIYIPGYKYKIVDGLITNHLPKSTLVMLVSALSSEDNLNAYKIAVENRYRFFSF
GDAMPK"
/complement(7305..8354)
/gene="CAC2284"
/feature="ruvB"
/complement(7305..8354)
/gene="CAC2284"
/transl_table=11
/product="Holliday junction specific DNA helicase, subunit
ruvB"
/protein_id="AAK80241.1"
/db_xref="GI:15025289"
/transl_table="MDDRILTSVNLEDSAYNRPQKLNIEYIGSKVKERMDIFIRA
AKKGEALDHLVFGPGLGKTLIANIAKEMGNLKVITSGPATIERAGDAAIITGLS
EYDLFTDEIHLRNSVEEILYPAMEDYALDIIIGKGAASKIRLDLPKFTLIGATFR
VGLLTAPLRDFGLCPMEFYNEDELKEIIVRSSILDIDIDEADALEIAMSRSRGTFR
IANLLKRVRYADVCKRGIYNLDSAKNALNLEVDDEGPDSDNKIIQAIYNNFAGS
PVGIETLSYFVGEIDTIEDVYEPYLLQKGFIMETPRGRMATKAYEHLKVPFNKKQ
GTVANDQCSFFKKEK"
/complement(8367..8972)
/gene="CAC2285"
/feature="ruvA"
/complement(8367..8972)
/gene="CAC2285"
/transl_table=11
/product="Holliday junction specific DNA helicase, subunit
ruvA"
/protein_id="AAK80242.1"
/db_xref="GI:15025290"
/transl_table="MYEYIKGYIGINKDYIIVDNGVGYYKIFTSGMTMANMPKPEED
VLIYLTQIVRQDFGLGYFLTKOBELEMFNKLITINGVAKASLSLLSICTVNNLKAYI

```

```

ITGDEKTIKAPGIGKKTQRIILILELKDPKFKSTDISKGNSEINNLDVVDYDEHSKKLEE
VRPALNSLGYSEKETDRAINNVNDKSEGIENIIKSLRFLMN"
/complement(9028..9675)
/gene="CAC2286"
/complement(9028..9675)
/gene="CAC2286"
/codon_start=1
/transl_table=11
/product="Uncharacterized protein, similar to protein from
Clostridium histolyticum (GI:3892648)"
/protein_id="AAK80243.1"
/db_xref="GI:15025291"
/transl_table="MTLKKVILSVVALLVTFIPSCIGYGRNLKIMNAEKYKISIG
IDEFLGRNESEKARYTSNNKGGKTLALLIVKQNYDITFDPEDETKNNVLTSSBETNA
EKEHKAGMTKEQITRSYEQGYIYIWNNDMKVEVIKNNYINNHNKVDQGVFLNKDKNSD
KINVYKVDNNGSLIKTNEQAYISVNQLKNETRLLRLLKGFVPGSKEDAKKLGYD"
/complement(9880..11235)
/gene="CAC2287"
/complement(9880..11235)
/gene="CAC2287"
/codon_start=1
/transl_table=11
/product="AcyL-CoA reductase LuxC"
/protein_id="AAK80244.1"
/db_xref="GI:15025292"
/transl_table="MKLEFTCGSEITIKENYKLEMLTSGKGDVEKELERLKNKNK
VHDIKTEETIDLLDRCAKLWDDNYGKKHETLQITNQSFELVNTVEFKSMOMLLRE
NIQNTIKRELGEVDILDSHWKNTYGYEHRQPRGIIFHNISGNAPVVPIMISMOLLK

```

Query Match 6.8%; Score 127; DB 1; Length 12592;
Best local Similarity 50.5%; Pred. No. 5.9e-18;
Matches 335; Conservative 0; Mismatches 325; Indels 3; Gaps 1;

Qy	198	TAAATCTGGAGCGATGCTTTTCTCCGAGGTTCTCAGTGAAGACGATCTCTCTCG	257
Db	1589	TAACTCAGGAGCTCTCTCTGTAACATAAAAGTTGTTCAATCCAAAACCTAGTAGTGCCT	1530
Qy	258	ATCTTGGGAAAAACAATGTACACAAAGGCATTCTCAGCATGCTGTGGCTTGGCAATGC	317
Db	1529	CACITGGTGAAGTGCACTTCCAAACAGTATACCTCAGGAGCAGTTGCTATATGTATAA	1470
Qy	318	TTATGTTTTGATGAGCGTATATATAGATTGGAGGGGTGATCGCTTCGGGAGCTGTTTC	377
Db	1469	TATCTTATTATCATGATTTAGTATATAGATACCTGCTTATGGCAGATATAGACTAG	1410
Qy	378	TTCTGAATCTTTTCTTATCTGGCAGCTCTACAGTATTTGGATGCGCCACTCACCTTGT	437
Db	1409	TTTATTATTACTGTTTGGTTTGTAGTGCAATTTAGCTGTAAAGTTTACATTAACACTTT	1350
Qy	438	CAGCACTCGCTGGGATTTCTTCTCTATGGGATGGCGGTAGATGCAAAATGTTCTTGAT	497
Db	1349	CAGTATAGCTGGACTTTTACTTACGATAGTATGGCTGTAGATGCCAATGTTCTTATAT	1290
Qy	498	TCGAAGAATCCGAGAGAAATTTTATTTGCTCAAAGTCTTAAAAAATCTCTAGAAAAAG	557
Db	1289	TTGAAGATTTTAAAGAAGAACTTAAAAATGGTAAAAACAGTTTAGATCAGCTTTTGATG	1230
Qy	558	GATATACCAAGCTTTTGGAGCCATTTTGTATCTTAATCTTAATCTACAGTATTCGCTCAG	617
Db	1229	GTTTTCATAGGGCTATGTCTCATATAATAGATTAATGTTTACCACCTATTATTTCAGGTG	1170
Qy	618	CACCTTCTTCTTCTCTAGATACAGGGCTATTAAAAAGGTTTGTCTTGACATTTGATTTAG	677
Db	1169	TTATACITTTATGCTCTTGGAGAGTGGCGAGGTAAAGGATTTGCGTTAACACTGGTAA	1110
Qy	678	GAATTTTCTCTTCAATGTTTACGGCTCTTTTTCATGACTAAATTTTTTTCTCATGTGGA	737
Db	1109	GTGTTCTTATAAGTATGGTACTGCTATTTATTTAACTAGAAAGACTTTTGTATGTTGGG	1050
Qy	738	TGATATAGACCC---AACATACACAGTTTGCATATGATGAATAAGTTCGTGGGGATAAGC	794
Db	1049	TAGCATGGGACTTTATAATAAAGCATCTCATTTCCGAGTTTAATCTCGAAGGAGTGATA	990

QY 795 ATGATTTCTTGAGAGATGCAAAACCTTGGCGCTTTCGSAAGTTTCTTTAG 854
 DB 989 AGTATGCTTAAGGTTTATAACACTTAAGTTGGTGGTATCTTACTATAAT 930
 QY 855 GTT 857
 DB 929 ATT 927

RESULT 12
 AP005369/c 300600 bp DNA linear BCT 17-AUG-2002
 LOCUS AP005369.1 GI:22293721
 DEFINITION Thermosynechococcus elongatus BP-1 DNA, complete genome, section 1/9.
 ACCESSION AP005369 BA000039
 VERSION AP005369.1 GI:22293721
 KEYWORDS
 SOURCE Thermosynechococcus elongatus BP-1 (strain:BP-1) DNA.
 ORGANISM Bacteria; Cyanobacteria; Chroococcales; Thermosynechococcus.
 REFERENCE
 AUTHORS 1 Nakamura, Y., Kaneko, T., Sato, S., Ikeuchi, M., Katoh, H., Sasamoto, S.,
 Watanabe, A., Iriyuchi, M., Kawashima, K., Kimura, T., Kishida, Y.,
 Kiyokawa, C., Kohara, M., Matsumoto, M., Matsuno, A., Nakazaki, N.,
 Shimpo, S., Sugimoto, M., Takeuchi, C., Yamada, M. and Tabata, S.
 TITLE Complete genome structure of the thermophilic cyanobacterium
 Thermosynechococcus elongatus BP-1
 JOURNAL DNA Res. (2002) In press
 REFERENCE 2 (bases 1 to 306600)
 AUTHORS Kaneko, T.
 TITLE Direct Submission
 JOURNAL Submitted (05-JUN-2002) Takakazu Kaneko, Kazusa DNA Research
 REFERENCE Institute, The First Laboratory for Plant Gene Research, 2-6-7
 AUTHORS Kazusa-kametar, Kisarazu, Chiba 294-0812, Japan
 JOURNAL (E-mail: kaneko@kazusa.or.jp,
 URL: http://www.kazusa.or.jp/cyano/Thermo/,
 Tel: 81-438-52-3935 (ex. 2338), Fax: 81-438-52-3934)

FEATURES
 source
 1. 300600
 /organism="Thermosynechococcus elongatus BP-1"
 /strain="BP-1"
 /db_xref="taxon:197221"
 /note="BP-1"
 129. 1142
 /gene="t1r0001"
 129. 1142
 /gene="t1r0001"
 /note="ORF ID: t1r0001"
 probable oxidoreductase"
 /codon_start=1
 /transl_table=1
 /protein_id="BAC07554.1"
 /db_xref="GI:22293722"
 /translation="MKAVATABEGEPDVLVROVPEPQIEADEVKIQLPASVNPID
 TKLRQRTFPDRPRALIGDGGAGVAVGAARPRVGEVFCVGLDRGGVAB
 YAVPEAAVAHAKPTLSFIQAAALPLAVITAMEALGDRGAVPPLNUTSTKTVLIHAG
 AGGVGHIAIQARRAGQAVATTISPPAKQFTALGTLAINITTDVAVLMDTGS
 KGVLDLADIVGATFAETPKAVRPGYGLATLLEPAVDTPMKIARQRLILQTLMLTP
 QLGLEAALAHQGIILEQVATLVDRGELRVVDKTFPLGAADAHRYLSQRLVQKGV
 Lip"
 complement (1351. 2331)
 /gene="t110002"
 complement (1351. 2331)
 /gene="t110002"
 /note="ORF ID: t110002"
 /codon_start=1
 /transl_table=1
 /product="serine proteinase"
 /protein_id="BAC07555.1"
 /db_xref="GI:22293723"
 /translation="MVABAGAVVSIPTLRDRGSEDPPLAPFPVDPVPLRQGGSGF
 IFTPDGKIMTAHYVEGASAVRYTLPGRODGVKLGADSLDVAVQINAKNLPYQ
 LGNSDTLPGEWATAIGNPLGSLNTVTAGIISANGRASSELGAADKXVSFIQTDALN

gene
 CDS
 complement (2402. 2578)
 /gene="t110003"
 complement (2402. 2578)
 /gene="t110003"
 /note="ORF ID: t110003"
 unknown protein"
 /codon_start=1
 /transl_table=1
 /protein_id="BAC07556.1"
 /db_xref="GI:22293724"
 /translation="MAADVQRNHVFQNKARVDLMSWVEYGPQAAEPMDGDRPWRPN
 GLSIFARSHPLP"
 2577. 2762
 /gene="t1r0004"
 2577. 2762
 /gene="t1r0004"
 /note="ORF ID: t1r0004"
 unknown protein"
 /codon_start=1
 /transl_table=1
 /protein_id="BAC07557.1"
 /db_xref="GI:22293725"
 /translation="WLTIVISYLEQPMFTFDSFFRGVTLQPGRNENVDERRMRCKTHNA
 DIALMKGLIYVEELG"
 2816. 4534
 /gene="urec"
 2816. 4534
 /gene="urec"
 /note="ORF ID: t1r0005"
 /codon_start=1
 /transl_table=1
 /product="urease alpha subunit"
 /protein_id="BAC07558.1"
 /db_xref="GI:22293726"
 /translation="MSYRIDRQTVAEYGPVIGDRLRLADNLIIEIHDYTHGDEV
 KRGGKVRIDMGQSPRIANABEAVVITNNVILPMKGVKADGCIKGGIKYIKGAG
 NPYTGEGVDIILIGPTEAIGAGMTLITNGGIDAHITFCPOQIATAIAGITTMIGG
 TGPATGTATTCIPGPWNIYRMLQADAFPVNLGFLGKNSSQPGQLIEQVAGVGL
 KLHEDMGSTPNAIDITCLSAEDYDIQVAIHDTVLNDSGFEVDITIAFNKRTIHAHVE
 GAGGGHAPDIKLCGOANVLPSTNPTPTRYVTNIDELMDLMVCHLDPISIPEDVAF
 AERSRIARETIAEDTLDHGARSIISSDSQAMRGESITRTWTAHMKVKGRLAD
 PORPGVDHNPRAKRYAKYTINPAITHTGIAEYVSVEGKADCLKRPAPFGKPE
 LVIKGVLAIVQMGDANASTPTPOPHMQPMFASGGCTTTSVTFMSQAGIANNIPB
 QLGKQTVLPVGHIRQLRKADKLNDYLPHEIVDETYEVADGELLTCEPATVPLA
 QRYFLP"
 4538. 5032
 /gene="t1r0006"
 4538. 5032
 /gene="t1r0006"
 /note="ORF ID: t1r0006"
 hypothetical protein"
 /codon_start=1
 /transl_table=1
 /protein_id="BAC07559.1"
 /db_xref="GI:22293727"
 /translation="MTDAVSALFQGGQQAFFRGKNTQOSITLLAQAVALAEGNTTQSGE
 ISLMLVTASAGDQEAVALSLCROLQRHDPDHTROBESRLILALEPOLKRRPWCSE
 IPDLSHLGDRAVTSYPRHSKRKPSAFTPKPQPPPATPLPNAFTWIALTGLGVATLL
 AAWL"
 complement (5430. 7640)
 /gene="t110007"
 complement (5430. 7640)
 /gene="t110007"
 /note="ORF ID: t110007"
 /codon_start=1
 /transl_table=1
 /product="cellulose synthase"
 /protein_id="BAC07560.1"
 /db_xref="GI:22293728"
 /translation="MSHEPPRPASPRQWAPMWVGFPVPSILLFGLVLLATTWRPTLF

```
PIDLVNPDQLRPLDLLQMEDELSVWMLVLVGVMAIALNFIPNNVSRLLIRFIVL
LFGRYLVWGVGCVSDIIFPIYNEPPYILRRTIVACQAIHYRNKSIYVLDGRSE
ANAYQQYALVGEVCPYDIIFPIYNEPPYILRRTIVACQAIHYRNKSIYVLDGRSE
IADICAGLVNLTLPNEHKNALHKTGELIADVADADIFPQNFILRTVGF
FQDDQSVMTQPHFNPVHSONLGIENPMGDMYFFGFIOGRDFGNALICCGTS
VWVRRALEAGVYTCVVEDFOTGFKMOIAGVRLVLYLNEILSMGESPFODYLEQ
RLRWLQNMQIYFCGDDLPVWSLWFSQRSCHLSLLLNINPIRFTCLGPRLSMT
GISULYALTELYALPULITLMIATFSWATBGRYFSLWGEVEVTFPGMVOLIKI
LRNPFGLISVNTKGLSNRKNRLRYTWPLVAFIAVSGVPIFYGGYWLHIWPM
EYDSEGLVMAWTLVNAFTALIAVLSIDQPSRRQDRFPVCTVCRFQGLDQCLYLR
TRVSETGAALLTAGRFISAORDTVGTLFLEODFSVTAAVRSRTEEBQCCLYLR
LEVSDEAQRGLVQLLYGLTWMHKPKAPNGLDAPWMLIRLDFERSLSLSYSH"
complement (7667. .8746)
/gene="ycf81"
complement (7667. .8746)
/gene="ycf81"
/note="ORF ID:tl10008
hypothetical protein"
/codon_start=1
/transl_table=1
/protein_id="BAC07561.1"
/db_xref="GI:22293729"
/transl_table="MLDTILLIIIVMGIAVGNSIDLLPETVLAQVANVRGLQWMA
AFGAIVGIALGLLOSLYHLERSIROLPELTLSRAVLGVGLLANLMLAPILFLP
IPKDFSIKPLIAVLTSILFAYSSTLLADSHGPALLRLINPNVASSILAEGLMKPAR
AKVLDTSICIDGRLEALNGLVLEGOIIVFQVLEQLQIADAGNEQKMRGRGLDV
LNRLQASIGDRIVHSDYPELTVDAKLVRLEOENGLTVNLDNLNKVARTFKQDV
FVNLELAQALRPYLPDGTLELKLKEGKEPAQGVGYLEDGTWVVVEGVHDIGDQLS
VWVTSALQTSAGRMIFARLOMPTWA"
complement (8796. .9758)
/gene="tl10009"
complement (8796. .9758)
/gene="tl10009"
/note="ORF ID:tl10009
unknown protein"
/codon_start=1
/transl_table=1
/protein_id="BAC07562.1"
/db_xref="GI:22293730"
/transl_table="MAETAPTAIAELVHTVGTDLRLRIQLKTDVGRDLSLTAYLKT
RGIOVHNPLAASITIEYHREQITPLQLLAAIQFMGEVOIIGQNGRGLONLRAFDL
EPREVGNKLTAMGFLGVLGDLIGVMGVTAGGLLGMGGVAMGVQVTFVGGVIG
ARLGEVYQNMQLQFALTEPPIAKALEIHSQDKIGGAAGELAGLAGOVVLGV
GETLGRVYNNVGSQLEDIGRLAEPSPGPPPSLHFFLEWMVKTSTQTFAKETLLA
TLGIVARGILGPOAEAGIRAGSRMRHLDRQOSNTAAKEKY"
9717. .10886
/gene="tlr0010"
9717. .10886
/gene="tlr0010"
/note="ORF ID:tlr0010"
/codon_start=1
/transl_table=1
/product="ABC transporter ATP-binding protein"
/protein_id="BAC07563.1"
/db_xref="GI:22293731"
/transl_table="MHELGNRRSFSGHNPQSQALVAVVARTPEGEISVQMLRSH
TELMTMTATVYHLRSYGAITAVADVSLTAHQEIGFVLGPNFGSKTTLRLCLTL
SRPDAGLEVCFVSHPLHROKLVGVAQVVALDKILTGQFLEQLAALYHPRAL
IPERIEAVLDRLQWRDROCGTYSGIRKRPDLAAGLHQPOVLVLDPTVGLDIE
SRQVINDVRLDKAQGLTILTSYHLEVDLLSRLVILDKGRVIAAGSPPEELKANIG
RDVTLVRREFTPQEAQRQALLUSQTVKASALINANQNSLNLVDTVDVTAATK
RALFEAGLPFLPSLAQSPSLDDVYLAATGTQLLDLALAAQAQRDSKOLKEVMQR"
10886. .11296
/gene="tlr0011"
10886. .11296
```

```
Query Match 6.7%; Score 124; DB 1; Length 300600;
Best Local Similarity 51.2%; Pred. No. 1.9e-17;
Matches 289; Conservative 0; Mismatches 275; Indels 0; Gaps 0;

Qy 142 TGCCAGTCTCTCAGGAATTTACCCACCGTCAGTGAGCAACTCGCCTCAGATTTAA 201
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 183759 TGCCGAATCTCCGGTGGCTTCACTGCCCAAACCGCCCAATGATCTGCCATTCAGTTGCA 183700
```

```
Qy 202 ATCTGGAGCGATGCTTTTGTTCCTCCGAGGTTCTCAGTGAAGAGACGATCTCTTCTGATCT 261
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 183699 GGGGGAGCGCTGCCGTTCCCTTAGAGGTGGTGGAAAAACCGACCGTGGTGTCTCCCT 183640
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 262 TGGAAAAACAATGTACACAGGCATTATCTCAGCATGCTGTGGCTTGGCAATGCTTAT 321
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 183639 GGGGCAAGACAGTATTCGCGATAGCCTCTATGCGGGCTTGGCCGGCTTGGGTGGTGTCT 183580
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 322 TGTTCATGAGCGCTATATATAGATTGGAGGCGTCATCGCTTCGGGAGCGTGTCTTCT 381
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 183579 AGTCTTATGTTGGCTTACTATCGCTCCCGGACTCATTTGCCGATATTGCCCTGCTTCT 183520
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 382 GAATCTTTTCTTATCTGGGAGCTCTACAGTATTTGGATGGGCCACTCACTTGTGTGAGG 441
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 183519 ATACGCCATTTTACCTATGCGCTCCCGGACTCATTTGCCGATATTGCCCTGCTTCT 183460
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 442 ACTCGCTGGGATTTCTTCTGCTATGGGATGGCGGTAGATGCAATCTTCTTGTGATTGCA 501
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 183459 GATTGGCGGTTTTTATCTCAGTATTGGCATGGCGGTGATGCCAACGTCCTGATTTTGA 183400
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 502 AAGAATCCGAGAGGAATTTTATTTGTCTCAAGTCTTAAAAAATCTCTAGAAAAAGGATA 561
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 183399 GCGACTTCGGAGAGATTACGGGACGGAACCTCTACGTTCCGTTGNAAGCGGAT 183340
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 562 TACCAAGCTTTTGGAGCCATTTTGTATCTTACTTCACTACAGTATTGGCCTCAGCACT 621
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 183339 CGATCGCGCTTTGCCAGTATTCTGGATAGCAACGTCACCACCTTGATTTGCTGTGTCG 183280
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 622 TCCTTTCTCTAGATACAGGCGCTATTAAAGGTTTGTCTTTGACATGATTTTAGGAAT 681
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 183279 CCTTTTTCCTTGGGACAGGTTTGTCCGCGGTTTGGCGGTACCCCTTGCATTTGGTAT 183220
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 682 TTTCTCTCAATGTTTACGCGCTCT 705
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 183219 TGGCATTAGTATTGTACGCTCT 183196
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

RESULT 13

```
AY078352 1557 bp DNA linear BCT 04-MAR-2002
LOCUS Rickettsia typhi export membrane protein SecD gene, complete cds.
DEFINITION Rickettsia typhi export membrane protein SecD gene, complete cds.
ACCESSION AY078352
VERSION AY078352.1 GI:19110413
KEYWORDS Rickettsia typhi.
SOURCE Rickettsia typhi.
ORGANISM Rickettsia typhi.
```

```
REFERENCE 1 (bases 1 to 1557)
AUTHORS On'gele,E.A., Radulovic,S. and Azad,A.F.
TITLE Characterization of the Sec operon of the typhus group Rickettsia
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1557)
AUTHORS On'gele,E.A., Radulovic,S. and Azad,A.F.
TITLE Direct Submission
JOURNAL Submitted (08-FEB-2002) Microbiology and Immunology, University of
Maryland School of Medicine, 655 West Baltimore Street, Baltimore,
MD 21201, USA
```

FEATURES

```
source
1. 1557
/organism="Rickettsia typhi"
/db_xref="taxon:785"
1. 1557
/codon_start=1
/transl_table=1
/product="export membrane protein SecD"
/protein_id="AAL82790.1"
/db_xref="GI:19110414"
/transl_table="MONLKKWIVPSIICVTFATICALPNFIOINMKFLPHDSVNLGL
DLRGGANLLDVDFDYLNDSDMENIADTRKPFREHKIGYKNLLVRHNNIQLVRSTE
VLSLKKIILHKIDPEIIIEVNNKKIKLYRSESLNDLLNKKVQESIEIVMRIDSTGT
KEPILOQGNKHTLLQVPGVDNPSLYLNILGKTAKLTFHLVDENANIEEAIKGHIPLD
```



```
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAC65400.1"
/db_xref="GI:3322696"
GFDHAAVVPVEEMQKFLQGNFSLDFLEKNKRLHROARSLHTEGKTVIRKK"
complement(3573..5471)
/gene="TP0413"
CDS
/gene="TP0413"
/notes="similar to GB:AE000783 percent identity: 35.60;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="phosphoglucomutase"
/protein_id="AAC65401.1"
/db_xref="GI:3322697"
VTLTKQADAVLSAQVALSYARFLKKVPKRVVIARSRPTGEIISICTAILHLEG
SLRCVGVPIPEIATYARNDAFLIYSASHNPFIYNGFKFLGNDGVLSPTQAAQSE
DFAQCQPASPIKIREHCATLNASILEPLLYNKKHQSLSHYDLFIREVIAQSPKE
TEHRAFFRTQRAAYSEQOTPLSILIDFNGSARAASIDRRLLSGLVALFSIAETP
GDIRHIVPSSSLTCAQAALTEAATRGSSPEERSIAFGFVDCDGRGNIVYDQTL
NRAVPHQAVFALSVSSEICELSRNLRMQPHRAPPIALVNGPISLRVEAIAQLLDV
HYFTEVGEAHLIEKAHLRLKRGVCYRILGEGSNGSILHPAAVRDPLHTVFALLKL
LLRGDQKQPLGFWCAHSHQTERQAHAFTISDLFASLPAYTTPTHTDTRALLQIRTT
DHALLKAYQRVFERRWAQEKDILKKRFGICSYRPILYRGREQDQAGDFSRADGGL
KIQYNGHQAVAFVPMVRGSKTEPIFRISADLKGNDGEAGEALLQLHADVMVREARCA
LSCADTTs"
5574..6929
/gene="TP0414"
CDS
5574..6929
/gene="TP0414"
PID:1220986 PID:1205130 percent identity: 43.24;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="D-alanine glycine permease (dagA)"
/protein_id="AAC65402.1"
/db_xref="GI:3322698"
/transl_table="MGVENVGEMHVAHVSFNVVLLGTGCVTVCMKFPFVRLM
YVLKQITGGRRGGKGSVSAFRAVSTAALATLGSNIVGVATAIAGGPGAFWIW
VTGIFGNTQFVSVLAVYRRDQDGRFVGGPMYLDGVPVGGVGLASLFCIFS
IASFGIGMTQSVNVALVDFDVCDDVRVTGAVLMVLGLVSVGLKSI5WVGMVVP
GMALYYCVGVVLVCCNTRQYLVSGAPAGTAAGVGFAGSVVRQIAIAGVIGRG
VAVNEAGLGTAPIAHAAITDHPVRQGLGFIFFVGTVMVSVTAFAILITGWQGG
ASGNALYQSFERGLPGICQYVGVNSVPLFAFTMLGWLIVGERCAFLCGGRVCP
YRAWLFPVYVGVAGVSSVWNISDAFNGLMALPNLVGLLFLARQGMRLREFFPAQOE
G"
6910..7443
/gene="TP0415"
CDS
6910..7443
/gene="TP0415"
/notes="hypothetical protein; identified by Glimmer;
putative"
/codon_start=1
/transl_table=11
/product="T. pallidum predicted coding region TP0415"
/protein_id="AAC65407.1"
/db_xref="GI:3322703"
/transl_table="MRSRGRGARRMWSHENGVAHGNVGVGAKFLSRCVLEKAE
EGEICWGVIVKVEQHDIRVKGEOHGFACGMWGDVGTGEIRAGDRTEEAHVTK
GGKVFRRQGVGGTDLRVLEKCRIVGVGMVGGPVVRVFCARSAHSAKRAQAY
LAWMPSIGCISPRVCPA"
complement(7390..8748)
/gene="TP0416"
CDS
complement(7390..8748)
/gene="TP0416"
/notes="similar to GB:AE000783 percent identity: 45.01;
identified by sequence similarity; putative"
/codon_start=1
```

```
/transl_table=11
/product="signal recognition particle protein (ffh)"
/protein_id="AAC65403.1"
/db_xref="GI:3322699"
/transl_table="MFEQLSASFKRIVGALGGRTATISERNIOEAVESIKRALDADVH
VVRVFNQVNTIQHAGQOTVLASVPAQOFIKIVHERLTAFLGEHRSLSHLKGPDTQS
IILLGLQSGSKTTSARLAAYLKADGRSPLLAACHDVRRAASAQLAVLGHVGPVY
QHALLPHEQPPCALDTAGQYARSHGNDVLIIDTAGRUHVDAALWOELLILLKTLVVP
VETLVADALATGQTVVRIABEFAAVGIGSVLWSKFSDTRGGAALSLSKITCOPLL
VTGGERPODFEPFHPERAAGRIILMGDIIVSLVEKAQKAFDAQSHAAKKTQSHERT
LSMDLHLOTIEKMGPLSLVEMIPLGLAVASADALDARAFKROKAIQOSMTVOERDN
FLIIGPSRRRIAGSGTSVADVNRILIKNFERMTLMRTAWQSRRAHPKGDTPYGWP
HR"
complement(8741..10420)
/gene="TP0417"
CDS
complement(8741..10420)
/gene="TP0417"
/notes="similar to GP:1651274 percent identity: 29.26;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="apolipoprotein N-acyltransferase (cutE)"
/protein_id="AAC65404.1"
/db_xref="GI:3322700"
/transl_table="MCTAPLVSSAASAVLLAFAINPEFWLAGSVLGLALVPLVYG
FLLSPAKHVACSYGLFVALVHACSSFNLKNFGFALFTLGASTVGFYFVALPFGVAF
ACTLRQAPARACAFALVWTLWVVKSTGLAYPWGTVPTMTAHSLSHLIQIADITGV
GSLFLPLANACVAESLHPIKRDSDVPVPRMLWLTGCLYCLSLYGAIVATLGAAPR
TLLALAIQONADPMDTXXSEKNTTAAIHLTETALOTAPPLTPPYRKEKTLTHAS
ARAPDMVWSESSLAYPYEQYRHYVNALPAARPSAFILTLGALLVGTPLRLSGNS
TKGGYANAVALLURPDGVAQVQGMQVFAEFIPWGHMTSVORLAQMLAGFSSEWTP
GPGPRLPHVPCAGGSVRPATPCYEDAPPSLCAALHTQGSSELLINDNSWKSATA
EMQYVYVSLFRATIELRTLLVSTNSGYTVVIGPEKTRAAPLFOATSAVLHVVPV
VRTYARMRDWVILCALIFFAEGVRMAVHTRHSTTQAESSLQIQIRGEHV"
10447..11640
/gene="TP0418"
CDS
10447..11640
/gene="TP0418"
/notes="similar to GB:U26401 SP:P51570 PID:1002507
PID:1929895 percent identity: 30.73; identified by
sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="galactokinase"
/protein_id="AAC65405.1"
Query Match 6.5%; Score 122; DB 1; Length 12504;
Best Local Similarity 56.1%; Pred. No. 7.6e-17;
Matches 230; Conservative 0; Mismatches 180; Indels 0; Gaps 0;
Qy 302 TGTGGCTTGGCAATGCTTATTGTTGATGAGCGTATATTATAGATTGGAGCGTCATC 361
Db 1423 TGGGGCTGTGCGCGGTGCTCTTATTATCTGCTGGTGTGATCAAGAAGCGGGCTAAC 1482
Qy 362 GCTTGGGAGACTGTTCTTCTGAATCTTTTCTTATCTGGGAGCTCTACAGTATTGGAT 421
Db 1483 GCTTGTGTCGCGAGCTTTTGAATCTGTATATCATGTTCGTGTCGCTTTCAGCGTTAA 1542
Qy 422 GCGCCACTCACTTGTGAGACTCGCTGGGATTTGTTCTTCTATGGGATGGCGGTAGAT 481
Db 1543 TTGACGCTTACGCTTTTCGAGTATCGCGGGGATGATTCCTCACTATTGTTGGCAGTGG 1602
Qy 482 GCAATGCTCTTGTATTGCGAAAGAAATCCGAGAGAAATTTTTTATGTCCTCAAGTCTTAA 541
Db 1603 GCGNATGCTGTTGCTTTGAGCGTATACGGGAAGAACTTGGCTGGGCAAAAGCCGGGG 1662
Qy 542 AAATCTCTAGAAAAGGATATACCAAGGCTTTTGGAGCCCAATTTTGTATCTTAACCTGACT 601
Db 1663 GCTGCTGTGTAGCGGCTTTGAGCGGTGCGTTTGGGCAATTTATGGATTCAACGTCGACG 1722
Qy 602 ACAGTATTGGCCTCAGACATCTTTTCTTCTAGATACAGGGCCTATTAAAGGTTTGCT 661
Db 1723 ACGTTTATAGCAGCGCTTTTCTCTTTCGGTGTCTCGGTACCGGTCTCTATTAAAGGGTTTCGCA 1782
```

QY 662 TTGCATTGATTTAGCAATTTCTCTTCATGTTAGCGCTTTTCAT 711
 Db 1783 TACAGTTGGCTATCGGGGATGCTGTCCTCGATTATTACGGCATGTTTCGT 1832
 RESULT 15
 RPKX03
 LOCUS
 DEFINITION Rickettsia prowazekii strain Madrid E, complete genome; segment 3/4.
 ACCESSION AJ235272 AJ235269
 VERSION AJ235272.1 GI:3861033
 KEYWORDS complete genome.
 SOURCE Rickettsia prowazekii.
 ORGANISM Rickettsia prowazekii.
 Bacteria; Proteobacteria; alpha subdivision; Rickettsiales; Rickettsiaceae; Rickettsiae; Rickettsia; typhus group.
 REFERENCE
 AUTHORS 1 (bases 1 to 279110)
 Andersson, S.G., Zomorodipour, A., Andersson, J.O., Sichenitz-Ponten, C., Alsmark, U.C., Podowski, R.M., Naslund, A.K., Eriksson, A.S., Winkler, H.H. and Kurland, C.G.
 The genome sequence of Rickettsia prowazekii and the origin of mitochondria
 JOURNAL Nature 386 (6707) 133-140 (1998)
 MEDLINE 99039499
 PUBMED 9823893
 REFERENCE 2 (bases 1 to 279110)
 AUTHORS Andersson, S.G.E.
 TITLE Direct Submission
 JOURNAL Submitted (11-NOV-1998) S.G.E. Andersson, Siv.Andersson@molbio.uu.se, Dept. of Molecular Biology, University of Uppsala, Husargatan 3, Uppsala, S-751 24, SWEDEN
 FEATURES
 source
 1. .279110
 /organism="Rickettsia prowazekii"
 /strain="Madrid E"
 /db_xref="taxon:782"
 63. .1045
 /gene="RP479"
 63. .1045
 /gene="RP479"
 /codon_start=1
 /transl_table=1
 /product="OCTAPRENYL-DIPHOSPHATE SYNTHASE (ispB)"
 /protein_id="CA14934.1"
 /db_xref="GI:3861034"
 /translation="MNTIIVKIQNLKDEVTOLNDLIISCLSDAELEKVGKYLEAVG
 GKRIIRPLITITAKMPYKGNKHKLASAVEPIHAALTDVDVDSNTLRRPPTANV
 IWSKTSILVGDPLFSQSFKLMAVSGCIKAMNTAKASVITSSEVQVYKLNERRII
 TIDBYQOIVKSKTALFGLACVGAIIAEVDVRSKQVNGRLGTFIVYIDLDLDY
 LGSBKQVGNIGDPLEGKVTLPILFYHKLQEDQKQMLNMLKSDKRTDDPVKIRD
 LMKHAIYNETVYVLSLENEANNLANKIPQNIKYKYLFSIRFILYRSY"
 1065. 1138
 /gene="trRNA Arg (CCG)"
 /gene="trRNA Arg (CCG)"
 1065. 1138
 /gene="trRNA Arg (CCG)"
 /gene="trRNA Arg (CCG)"
 complement (1194. .1265)
 /gene="trRNA Gln (TTG)"
 complement (1194. .1265)
 /gene="trRNA Gln (TTG)"
 /gene="trRNA Gln (TTG)"
 complement (1316. .3091)
 /gene="RP482"
 complement (1316. .3091)
 /gene="RP482"
 /codon_start=1
 /transl_table=1
 /product="unknown"
 /protein_id="CA14935.1"
 /db_xref="GI:3861035"
 /translation="MRHDRIILRLKFLFLENIGEYIIPSNDKYNNEVPEAKLEFY
 TGTGSSGIALICDAAPFTPDGTYLEQANKLEDLAFKYTDLKEIYKTLQKNIKIGY
 DPQLFTYQVLANNINPHKINENLVDKIWNKPILEPNSKYIHLDIKAGVSHNDKIRK

CRETILSSSVTVGCKNNDDILVILDSASICMLNLRLASDVNTPLMEKAVILITSTK
 IYLFNIPRIDPEIILNABPEITILPEKEENILPISKRYILIDDSITSVHMDLANK
 KYKATVEPCLLAKACKNDIEIKHALIDFKDAVALCEPFAEFTLYHSNVCCHSH
 ELITHSILCLKLTAPAKOEGVSDSFNAICGFQNSAIHYRANPKAKKIEGSHIL
 LIDSAQYKATDIDYTRTIIIVGIPCEQKRYOVLKHAITLAKFPKNIVYGANLD
 ILARQYLMQDMIDYDHTGCHGVGSFLSVIEGPOSINLSKILKGMILSNPGFYIP
 GKYGRIENLIYKENNGMLPEFETSLVPYAKLIDVALNLINDEINYIKEYNKRIRAK
 IYNILSTQAKWNLNINFLQSL"
 complement (3174. .4463)
 /gene="RP483"
 complement (3174. .4463)
 /gene="RP483"
 /codon_start=1
 /transl_table=1
 /product="PUTRESCINE-ORNITHINE ANTIporter (pocE)"
 /protein_id="CA14936.1"
 /db_xref="GI:3861036"
 /translation="MSOKLSFWAVPALVTGSGTGFILPLSLAPFGVYSIMGILS
 LFGMSIALVFSCICTKPEKTPGPHVYRASGDTIAPFTGWTWIIISVTSIVIS
 ALGYLTPPKSOTIILDIQLILALAIATLNKSKIKAKVEFYTLTKFVPLAVGL
 ALVFNHINIVYIAKVENFTPTMGVALLTFMGFTIGIBCATTTAGVADPAKTIP
 RAIIITIGCCVAVLYINSIGIIGLIPASELINSKAPVADAAITLFGGWSKVITYAS
 VICIGTLNMAVLTSGOIALGLADEDLPEFKKNSNNAPTYGIIISCIQITPLLEFT
 SNNPAKQITQIIDFSVIAFLFVYLICSLAPLAKVIFSSKENSYYLFWAIIISIFT
 WVIYKTPPETLIACSFITLIGIPVYGYWRLNRIKL"
 complement (4633. .4965)
 /gene="RP484"
 complement (4633. .4965)
 /gene="RP484"
 /codon_start=1
 /transl_table=1
 /product="HESB PROTEIN (hesB)"
 /protein_id="CA14937.1"
 /db_xref="GI:3861037"
 /translation="MKNVISLTDAAKQVYKLIBRAXPTFGIRVINGSGCAGQTVY
 VEYADNKNQFDEVEEKGVRILIDPKLMIYILGSEMDYVETNFKSQFPTNERNANC
 GGKSGPSV"
 complement (4969. .5364)
 /gene="RP485"
 complement (4969. .5364)
 /gene="RP485"
 /codon_start=1
 /transl_table=1
 /product="NIFU PROTEIN (nifU)"
 /protein_id="CA14938.1"
 /db_xref="GI:3861038"
 /translation="MAYSKKVIDHYENPNVNSLDKKKNVGTGLVGAPACGDVAKLQ
 IGVGDEIITDAKFTFGCGSALASSLVTEWIKGSYEDAKEIKNTEIAKELSPV
 KHGSLIADAIKKAIAIDYKOREKKOS"
 complement (5430. .6662)
 /gene="RP486"
 complement (5430. .6662)
 /gene="RP486"
 /codon_start=1
 /transl_table=1
 /product="NIFS PROTEIN HOMOLOG (sp11)"
 /protein_id="CA14939.1"
 /db_xref="GI:3861039"
 /translation="MNOQLKNTLPIYMDYSTPDIIDPVMEAMLPYFTTKGNPNSR
 SHSPGEAENAVENARSMVAKYIGADSKELIPTSGETSNNIVIGIAKFGYNNKKHII
 TILVSEHKVNLACRHLBOEGIKITYIPKSGIIDLFTLKAITDGTILAVVMVNN
 EIGVIOPLKEIGKICRERNVFPNSDIAQFGKIPINANECDNLASISHKIYGRKGI
 GALVIRKPRVATPLINGGQGERKRSGLTLPILVGIISAEIAYNMEKOTQIVN
 YLDFRPLNNIHSKISEVYINGDKQKQVGNLMLSPAGVEGSIILAIKDLAVSSGAC
 TSASLEPSYVLSIGISEELATSIRFGIGRFTTQEDIDYAVNLVCSKIDKLRLSPL
 MEMMOGVDLKKIRWTAA"
 complement (6740. .7870)
 /gene="RP487"
 complement (6740. .7870)
 /gene="RP487"
 /codon_start=1
 /transl_table=1
 /product="NIFS PROTEIN HOMOLOG (sp11)"

AUTHORS Aoki, K., Oguchi, A., Hosoyama, A., Nagai, Y., Kuroda, M., Hiramatsu, K., and Kikuchi, H.
TITLE Direct Submission
JOURNAL Submitted (30-JAN-2001) Akio Oguchi, National Institute of Technology and Evaluation, Biotechnology Center; 2Chome 49-10 Nishihara, Shibuya-ku, Tokyo 151-0066, Japan
COMMENT (E-mail: oguchi@nite.go.jp, URL: http://www.bio.nite.go.jp/, Tel: 81-3-3481-8423; Fax: 81-3-3481-8424)
On Jun 12, 2001 this sequence version replaced gi:13701258.
FEATURES
source
1. .301550
Location/Qualifiers
/organism="Staphylococcus aureus subsp. aureus N315"
/strain="N315"
/sub_species="aureus N315"
/db_xref="taxon:158879"
/complement (249. .947)
/gene="SA1293"
/complement (249. .947)
/gene="SA1293"
/note="ORFID:SA1293"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAB42553.1"
/db_xref="GI:13701259"
/translation="MSITFTMIYFVYLMLPLMAQHKVKSNTKISQVNSTGKTGR
HVALEIHLANGIDVEVKGEGFLTDHDPKKVSLSPANYRPSVAGTALAHEVG
HA1ODHQGFLFRALVPVANGSLSTYIMLGIIITPAISAFGSTALMTGAGIM
SLAVFISITLPEVFSDASSRAMKQITALINVEKEYHAKVLSAAMTYVATAVAV
AELVRIILVARSDD"
/complement (1001. .1588)
/gene="SA1294"
/complement (1001. .1588)
/gene="SA1294"
/note="ORFID:SA1294"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAB42554.1"
/db_xref="GI:13701260"
/translation="MTINTFMQYTLVQSRWLMMLLGNILGMIYGIWYGEOLSHPTP
HPKIFVDPSPALVYLTISLILKONKSIIEALVFLPFTKGIWAVINILFIEQ
GDITVNGIYLMFSGSIMAVQAIYVYPRKSMGISVAMTWFLNDIDVFIHQFPYI
DFITTHWQJGVSSCCSLVFGLLIYIELNKLCKK"
/complement (1578. .2153)
/gene="SA1295"
/complement (1578. .2153)
/gene="SA1295"
/note="ORFID:SA1295"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAB42555.1"
/db_xref="GI:13701261"
/translation="MSETLNQIKESFIEYLLFOYRPSRIYAVWVINYIKYNEKLANI
HFDTKINHELEIEAVGSHASAIQTKKNIKAMNTNEIPDIYANHCAPFDIOHPA
NVSRRQRDLIVAQTESPSYQTYIAHNSAIDHKKALLIDVLIHNDISLQNN
EKORFYQIQLINTLKLKNKNOFEDLADD"
/complement (2167. .3411)
/gene="SA1296"
/complement (2167. .3411)
/gene="SA1296"
/note="ORFID:SA1296"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAB42556.1"
/db_xref="GI:13701262"
/translation="MEDIKYKLIIDINQKLENTLSRVNEAITTDNDALFILGETLYN
FGMPQGLIEVRVLYHKYDPSEELLIFIEGLMSENQTDALREYLSVSPSPKMLME
ADLYQOINMEVAIDKLOALELEBPNDIHPFALAEMLYDGOYLRSYSEYVLENG
EYONGNVLFSRMADCSLQSGNYSADRIYDEINDEMTSEDTYIKKAIISYDKNDIOE
AIKIMTTLISKPDYIQGYLYLQSLYENKKNYPDAIETGKEGLRLSQFYKELMTTTC

LEIEHGDANEGLLILKQALEVNDAYOEPLILISDLVYRNEDEYAEIITELTYDEEDLD
PFFPMWHLARAYQOERDKKQAEQFELAIPTMTKNTIDPNSDYIFYLEIGQKQAFITLL
NQLLEBPSNEMWDELRLQS"
/complement (3418. .4716)
/gene="aroA"
/complement (3418. .4716)
/gene="aroA"
/note="ORFID:SA1297"
/codon_start=1
/transl_table=11
/product="3-phosphoshikimate 1-carboxyvinyltransferase"
/protein_id="BAB42557.1"
/db_xref="GI:13701263"
/translation="MVSROIIDISGPLKGEIEVPDCKSTHRAIMLASAETSNIYK
PLUGEDCRIMDIFRLGVDIKEDDKLVNSPGYKAFKTPQVLITGSGITRLLA
GLLSGIGIESVLSDVSIGKRPMDVLRPLKMDANIGIEEDNYPLIIPKSVIRGIN
YQMEVASQOVKSAIILFASLFSNDITVIELDVSRRHTEFMFHNPIPLEASLSTTT
PDALDHKADFPVPGDISSAAFIVAAALITPESVTHNVGINPTRSGIDIVKMG
GNIOLEPNOTTGAEPTRASIRIOYTPMLQPTITGELVPKAIIDELPVIALICQAVGST
TKDAEBLKRYKETNRIDTTRDMNLNLGFELOPTNDGLIHPSEFKTNATYDSLTDRIG
MMLAVASLSSSEPVKIKOPDAVNSPQGLPRLKLLNENG"
/complement (4726. .5790)
/gene="aroB"
/complement (4726. .5790)
/gene="aroB"
/note="ORFID:SA1298"
/codon_start=1
/transl_table=11
/product="3-dehydroquinate synthase"
/protein_id="BAB42558.1"
/db_xref="GI:13701264"
/translation="MKLOTTYPSNNYPIFVEHGALDHSTYIDQDQSFILDEHYNO
YFADKFDLISLENVHKTIPAGEKTKTEBOYQETLEYILSHVTRKNTALVGGAT
GDPAGFVAATLLRGVFIQVPTTILAHDSVQKGINSKQKNLIGAFRPTAVIYD
LDFLKLTPPEQILSGYAEVYKHALNGESTTOEIQHFQDRILQSLNGMDKYAKGI
ETKLDIVADEKEQGVKRFNLNGHTEGHAVEYNHKILAGHAMVIGIYQIVANILEN
SNHDIQHYNYLTIXKGYPLETITDIDFTIYQYMLDKKNDQSGQWMLIKHFGDIIV
QHIDQTLQAHCEQLKTYRK"
/complement (5816. .6982)
/gene="aroC"
/complement (5816. .6982)
/gene="aroC"
/note="ORFID:SA1299"
/codon_start=1
/transl_table=11
/product="choistamate synthase"
/protein_id="BAB42559.1"
/db_xref="GI:13701265"
/translation="MRYLTSGESHPOLTVIEGIPANLEIKVEDINKEMFKROGGY
RRRMQIEKDTVEIYGVNNGYTLGSPITMYVTNDPTMRKMGAAPISEERENMK
RTITKPRGHADLVGMYKNHRDLKLVLEERSARETARAVAGLCKVLLQDLIDYI
SSVEIEGGIKDQFYDSFETFKANLDRNDRVIDSIAQMRKIDKARNEGSGISGV
QVVENMEPVGVSYYHYDRKLDKGIAQGVSNARFVSGFGFPAAEKPGSIODEI
LYNSEIGYRGSHLGLGEGSGNMPILVNGVMKPIPLTYPLNSVDINTDEPKAT
IRSDSCAVPASIVCEHVAREIKALLERQSHHIEQLQQQIADRQLANVF"
/complement (7368. .7568)
/gene="SA1300"
/complement (7368. .7568)
/gene="SA1300"
/note="ORFID:SA1300"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB42560.1"
/db_xref="GI:13701266"
/translation="MKNISITILGILMFLKLMELLYGAIFLDPKPLNPITKILFILILI
YIFVLVXELIIFLKSXYNKA"
/complement (7597. .7722)
/gene="SAS045"
/complement (7597. .7722)
/gene="SAS045"
/note="ORFID:SAS045"
/codon_start=1

	/transl_table=11	
	/product="hypothetical protein"	
	/protein_id="BAB42561.1"	
	/db xref="GI:13701267"	
	/translation="MWLIVIMIHKKARFLCCSKLEKFYSLVTRIKRGIFPIK"	
gene	complement(7777..8226)	
	/gene="ndk"	
CDS	complement(7777..8226)	
	/gene="ndk"	
	/note="ORFID:SAL301"	
	/codon_start=1	
	/transl_table=11	
	/product="nucleoside diphosphate kinase"	
	/protein_id="BAB42562.1"	
	/db xref="GI:13701268"	
	/translation="MERTFLMKPDAVQRNLGEVISIERKGKLGVGGKLMQVPMBEL AETHYGHGCKPFNDLIISFTAPFAVMVVEGDAAVVNSRHIIGSTNPSEASPGSIR GDLGLTVGRNIHGSDLSAESAREINLWFNENETTSVASPRDAWLYE"	
gene	complement(8318..9277)	
	/gene="gerCC"	
CDS	complement(8318..9277)	
	/gene="gerCC"	
	/note="ORFID:SAL302"	
	/codon_start=1	
	/transl_table=11	
	/product="heptaprenyl diphosphate synthase component II"	
	/protein_id="BAB42563.1"	
	/db xref="GI:13701269"	
	/translation="MAKLNNNEIKKVEORLEKAISKDVLQASILLSSCGKVR PAFVILSQSGKDIEQSEQYYQVAVALIELHMATLVHDDVIDKSRRRGLKITISKKW	

Query Match 6.4% Score 119.4; DB 1; Length 301550;

Best Local Similarity 51.0%; Pred. No. 2e-16;

Matches 282; Conservative 0; Mismatches 271; Indels 0; Gaps 0;

Qy	174	AAGTGAGCAAACTCGCCTCACATTTAAATCTGGAGCATGTCTTTTGTCCCAGGTTTC	233
Db	173176	AAGCAAAAACAAATCGCCGAATTGTTAAATGCAGGTTCCTTACCAGTTGCATTTGAAGAATA	173117
Qy	234	TCAGTGAAGACGATCTCTTCGTATCTGGAAAAACAATGTACACAGGCATTATCT	293
Db	173116	TTTTATCAAATCAGTAGTGCCCAATTTGGTCAAGATAGATTAATAACTGTGTGTTG	173057
Qy	294	CAGCATGCTGGCTTGGCAATGCTTATTTGTTGATGAGCGTATATPATAGATTTGGAG	353
Db	173056	CATCGTTATTGGTGTAGCATTAATTTATTATTATTCATCGTTGGATTCTACCGATTACCTG	172997
Qy	354	CGGTCATCGCTTCGGAGCTGTTCTTCGAATCTTTTCGCTTATCTGGCGACGCTCTACAGT	413
Db	172996	GTTTAGTAGCGAATTATTGTCATTTGACACTTATATCTATCTAACCGTTAGTACATTTAAT	172937
Qy	414	ATTTGGATGCCCACTCACTTGTTCAGACATCGCTGGGATTTGTTCTGCTATGGGATGG	473
Db	172936	TCAITTCGGGGTCTAACTTTACAGATTTAGCGCGTTCGGTATAGGTGTAGTATGG	172877
Qy	474	CCGTAGATGCAAAATGTTCTTGATTTCGAAAGAAATCCGAGAGAAATTTTATTTGCTCAAA	533
Db	172876	CTGTAGATGCCAATATTATCATGTATGAGCGTATTAAAGATGAATTCGAATAGTAGAA	172817
Qy	534	GTCTTAAAAAATCTGTAGAAAAAGATATACAGGCTTTTGGAGCCCAATTTTGAATCTA	593
Db	172816	CGATAAAGCAAGCCCTTTCTAAAGCAAAACAAAAGTTCAATCTCTAACAAATTTTGAATCTA	172757
Qy	594	ACTTGACTACAGTATGGCCCTCAGACATCTCTTTCTCTAGATACAGGGCCTATTAAAG	653
Db	172756	ACTTAACAACAGTTATCCCGCAGCAGTATTATCTCTTCGTTGAAAGTTTCAGTTAAAG	172697
Qy	654	GGTTTGCTTTGACATTTGATTTTAGAAATTTTCTCTTCAATGTTTACGGCTCTTTTCATGA	713
Db	172696	GTTTCGCGACAATGTTATTATTAGGTATCTAANTGATCTTTGTTACAGCGGTCTTAT	172637
Qy	714	CTAAATTTTCTT	726

Db 172636 CAAGATCTTATT 172624

RESULT 17
AP004827/c

LOCUS AP004827 33750 bp DNA linear BCT 02-JUL-2002

DEFINITION Staphylococcus aureus subsp. aureus MW2 DNA, complete genome, strain:MW2, section 6/10.

ACCESSION AP004827 BA000033

VERSION AP004827.1 GI:21204509

KEYWORDS

SOURCE Staphylococcus aureus subsp. aureus MW2 (strain:MW2) DNA.

ORGANISM Staphylococcus aureus subsp. aureus MW2

REFERENCE 1 Bacteria; Firmicutes; Bacillales; Staphylococcus.

AUTHORS Baba, T., Takeuchi, F., Kuroda, M., Yuzawa, H., Aoki, K., Oguchi, A., Nagai, Y., Iwama, N., Asano, K., Naimi, T., Kuroda, H., Cui, L., Yanamoto, K. and Hiramatsu, K.

TITLE Genome and virulence determinants of high virulence community-acquired MRSA

JOURNAL Lancet 359 (9320), 1819-1827 (2002)

MEDLINE 22040717

PUBMED 12044378

REFERENCE 2 (bases 1 to 33750)

AUTHORS Aoki, K., Oguchi, A., Nagai, Y., Asano, K., Iwama, N., Baba, T., Kuroda, M., Hiramatsu, K. and Kikuchi, H.

TITLE Submitted (06-MAR-2002) Akio Oguchi, National Institute of Technology and Evaluation, Biotechnology Center; 2Chome 49-10 Nishihara, Shibuya-ku, Tokyo 151-0066, Japan (E-mail:oguchi@nitech.go.jp, URL:http://www.bio.nite.go.jp/, Tel:81-3-3481-8423, Fax:81-3-3481-8424)

JOURNAL Location/Qualifiers

FEATURES source

1..33750

/organism="Staphylococcus aureus subsp. aureus MW2"

/strain="MW2"

/db_xref="taxon:196620"

complement(278)..619

/gene="MW1341"

complement(278)..619

/gene="MW1341"

/note="ORFID:MW1341"

/codon_start=1

/transl_table=11

/product="hypothetical protein"

/protein_id="BAB95206.1"

/db_xref="GI:21204510"

/translations="MTNITKEVDNLEQIDVFAKNKTGSSAKPYLDEYHSHKIDYFKVNDITGNIDPDLKQYVPVPMFKERYDYNIERYKHYFMGYRQMKTFKTELKMNNA SYOTRLKNKQV"

complement(624)..1283

/gene="nth"

complement(624)..1283

/gene="nth"

/note="ORFID:MW1342"

/codon_start=1

/transl_table=11

/product="endonuclease-like protein"

/protein_id="BAB95207.1"

/db_xref="GI:21204511"

/translations="MVSKKKALMIDVIANMFPDACELKNDHPFELTIIVLLSAQCTDVLNVTTELPKYKTPDYLVASDEELMNDIRISGLYRNKAKNTKLCQSLIDQFNGETPQTHKESLAGVGRKTAQNVMSVAFDEPSLAVTHYRVSKEGLNWRKDNVROVEDRLCSVIPDRNRNNSHQILIFFGRYHCLARKPKCDICPLEDDREGQKRYKASLKEA"

complement(1273)..1959

/gene="MW1343"

complement(1273)..1959

/gene="MW1343"

/note="ORFID:MW1343"

hypothetical protein, similar to chromosome replication initiation protein dnab

/codon_start=1

gene

CDS

gene

CDS

gene

CDS

gene

CDS

Query Match	Best Local Similarity	Mismatches	Indels	Gaps
Query 174	6.4%; Score 119.4; DB 1; Length 333750;	0;	0;	0;
DB 227352	AAGCAAAACAAATGCCGAATGTTAAATGACGAGTCTTTACCGAGTTCCTTGAAGAA	227293		
DB 227352	TTTATTCACATCTGAGTAGTGCCCAATTTGGTCAAGATGATTTAGATTAACATGTGTTG	227233		
DB 227292	TTCATGTAAGAGACAGATCTCTGATCTTTGGGAAAAACATGTACACAGCATTAATCT	293		
DB 227292	CAGCATGCTGTGGCAATGCTTAATGTTGTTGATGAGCGTAATTAATTAAGATTTGGAG	353		

Db 227232 CATCGTTATGGTGTAGCATTAATTTATTTATTTATTCATGCTTGGAATCTACCGATTACCTG 227173
Qy 354 GCCTCATCGCTTCGGGAGCTGTTCTTCTGAACTCTTTTGGCTTATCTCGGCGAGCTCTACAGT 413
Db 227172 GTTTAGTAGCGGATTTATGCAATTCGACAACTTATATCTATCTAACGTTAGTAGCATTTAATT 227113
Qy 414 ATTGGATGCCCACTCACCTTGTGAGGACTCGCTGGGATGTTCTTGGCTATGGGGATGG 473
Db 227112 TCATTTCCGGGGTTCTAACTTTACAGGATTAGCGGGTGGTATTAGGTGATGATGG 227053
Qy 474 CCCTAGATGCAAAATGTTCTTGTATTCCGAAAGATCCGAGAGGAATTTTATTGCTCTCAA 533
Db 227052 CTGTAGATGCCAATTTATCATGTATGAGCGTATTAGGATGAACCTTCGAATAGGTAGAA 226993
Qy 534 GTCTTAAAAATCTGTAGAAAAAGGATATACCAAGGCTTTTGGAGCCATTTTGAATCTA 593
Db 226992 CGATAAGCAAGCCTTTCTTAAAGCAAAAGAGTTCAATCTCTCAACAATTTTGAATCTA 226933
Qy 594 ACTTGACTACAGTATGGCTCAGCACTCTCTTTCTCTCTAGATACAGAGGCTATTAAAG 653
Db 226932 ACTTAACAACAGTATTCGCGCAGCAGTATTATTCTTCTCGGTGAAAGTTCAGTTAAAG 226873
Qy 654 GGTTCGTTTGACATTCGATTTTAGGAATTTCTCTTCAAGTTTACGGCTCTTTTCATGA 713
Db 226872 GTTTCGGCAAGATTATTATTAGGTATCTTAATGATCTTTGTTACAGCCGTGTTCTTAT 226813
Qy 714 CTAAATTTTCTT 726
Db 226812 CAAGATCTTATT 226800

RESULT 18
AP003363/c
LOCUS
DEFINITION
Staphylococcus aureus subsp. aureus Mu50 genomic DNA, complete
sequence, section 6/9.
ACCESSION
AP003363 BA000017
VERSION
AP003363.2 GI:14247399
KEYWORDS
SOURCE
Staphylococcus aureus subsp. aureus Mu50 (sub_species:aureus Mu50,
strain:Mu50) DNA.
ORGANISM
Staphylococcus aureus subsp. aureus Mu50
Bacteria; Firmicutes; Bacillales; Staphylococcus.
REFERENCE
1
AUTHORS
Kuroda,M., Ohta,T., Uchiyama,I., Baba,T., Yuzawa,H., Kobayashi,I.,
Cui,L., Oguchi,A., Aoki,K., Nagai,Y., Lian,J., Ito,T., Kanamori,M.,
Matsumaru,H., Maruyama,A., Murakami,H., Hosoyama,A., Kaito,C.,
Mizutani-Ui,Y., Takahashi,N.K., Sawano,T., Inoue,R., Kaito,C.,
Sekimizu,K., Hirakawa,H., Kuhara,S., Goto,S., Yabuzaki,J.,
Kanehisa,M., Yamashita,A., Oshima,K., Furuya,K., Yoshino,C.,
Shiba,T., Hattori,M., Ogasawara,N., Hayashi,H. and Hiramatsu,K.
Whole genome sequencing of methicillin-resistant Staphylococcus
aureus 357 (9264), 1225-1240 (2001)
JOURNAL
Lancet 357 (9264), 1225-1240 (2001)
MEDLINE
21311952
REFERENCE
2 (bases 1 to 342600)
Ohta,T.
TITLE
Direct Submission
JOURNAL
Submitted (28-FEB-2001) Toshiko Ohta, University of Tsukuba College
of Medical Technology and Nursing, Department of Medical
Technology; 1-1-1 Ten-nodai, Tsukuba, Ibaraki 305-8577, Japan
(E-mail:tohta@akura.cc.tsukuba.ac.jp, Tel:81-298-53-3454,
Fax:81-298-53-3454)
COMMENT
On May 29, 2001 this sequence version replaced gi:13875943.
FEATURES
Location/Qualifiers
source
1. .342600
/organism="Staphylococcus aureus subsp. aureus Mu50"
/strain="Mu50"
/sub_species="aureus Mu50"
/db_xref="taxon:158878"
complement (42. .815)
/gene="SAV1628"

CDS
complement (42. .815)
/gene="SAV1628"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAB57790.1"
/db_xref="GI:14247400"
/translation="MKHQSFRNELAIGQGLDLKQQTIVILGVGVGFAAEALART
NIGHTILDKDDVITNVNRQHALTTVQSKVSLMEERIKLINPDCKVTPLHMYFT
BETYEEDFNVDYDIFVDASDTIMYKVLMEKCLDRGIKVISMGAAKNTDPTFTIS
DKSTHTPMAKIIRQKLKLGIRKGIPIVPSDESPIVIREVDVKDVGKNAINRKGQ
MPPSNAFVPSVVGILICASYVNDVKDIPVRIRKDGQ"
complement (1029. .1130)
/gene="SAV1629"
complement (1029. .1130)
/gene="SAV1629"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB57791.1"
/db_xref="GI:14247401"
/translation="MGFLAAATPYMKDVQNKQKGTHLYITGRMIKLL"
complement (1276. .3042)
/gene="asps"
complement (1276. .3042)
/gene="asps"
/note="SAV1630"
/codon_start=1
/transl_table=11
/product="aspartyl-tRNA synthetase"
/protein_id="BAB57792.1"
/db_xref="GI:14247402"
/translation="MSKRTTYCGLVTEAFLGQEIITLKGWNNRRDGGILFVLDLRDRE
GIVQVFNPAFSEALKIAETVSEVVEVQGTVKRDPETVNPKIKTQGEVQVQVNI
KIVKSETPPSFINENNVNVDENIRKYRLDLRQELAQTFRHQITRSIRQYVLD
EGFPIETPLTKSTPEGARDYLVSFRHDEGFALPQSPQLFKOLLMSGDKYIQI
VKCFEDLRADKQPEFTQVDIEMSFVDQEDVMQGEEMLKVKVEKVGVEINGATPR
MTYKAMRRYSGDKPTFPEMLIDVSQLRDMDFKFDVTENDEGEIKAIIVAKGAE
QYTRKMDALTEFNIYGAKLARVKLAKELGIDETKLNFLVFTDWPVLLLEYDEDAKRYVAA
LVMPVADKPNVVAQSLGARVKAELGIDETKLNFLVFTDWPVLLLEYDEDAKRYVAA
HHPTSPKEDIKLGTPAEAEANAYDILANGVELGGGIRIHGELQEKMPFVLGF
TKEAQEQPGFLLDAFKYCAPPHGGIAGLGLRLVLLTNRNLRTIAPKPTASATCL
LTNAPGEVSDKQLESLSLRH"
complement (3058. .4320)
/gene="hiss"
complement (3058. .4320)
/gene="hiss"
/note="SAV1631"
/codon_start=1
/transl_table=11
/product="histidyl-tRNA synthetase"
/protein_id="BAB57793.1"
/db_xref="GI:14247403"
/translation="MIKIPRGQDILPDSKQWRYIENOLDMLTFYNYKEIRTPIFE
STDLFARGVSDTVQKEMYTFKDKGRSITLRPEGTAAVRSYIEHKMQGNPQPI
KLYNGPMFPRYERKQRFQNFQVEAIGAEINPSVDAVLAMVNIYOSFGLKHLK
LVINGVDMASKEYNEALVGHFEPVIEHFCDCQSLRHTNPMRILDCVKDRDKEAIK
TAPRITDPLNEESKAYEVOVKAVLDLGIPIYEDNLVRLGLDYTHTAPELMDNPNY
DGAITTLGGGRYNGLLELLDGPSETGIFALSIERLLALSEEGELDIEENLDLPI
VMTGQADRYAVKLNHLNHNKIGKADKYLQRKIKGQMKQADRUGAKFTIVIGDQBLE
NNKIDQKNMTTGESETIEDALVEYFKK"
complement (4782. .5657)
/gene="lyth"
complement (4782. .5657)
/gene="lyth"
/note="SAV1632"
/codon_start=1
/transl_table=11
/product="N-acetylmuramoyl-L-alanine amidase"
/protein_id="BAB57794.1"
/db_xref="GI:14247404"
/translation="MKKEAWLSKKGLKNKRTLIVVIAFVLPIFLFLLNSNSEDG
NITITENAEALRTGPNAAVPVIYKVEKGFHKIGKVGKWEVEDTSSNEKGIAGWHT

AUTHORS		TITLE		JOURNAL		FEATURES		BASE COUNT		ORIGIN		Query Match		Best Local Similarity		Matches		Indels		Gaps	
source		1. 45613		/organism="Moraxella catarrhalis"		/db xref="taxon:480"		13249 a		9346 c		9947 g		13071 t		6.3%; Score 118; DB 6; Length 45613;		50.5%; Pred. No. 5.1e-16;		0; Mismatches 280; Indels 0; Gaps 0;	
Legace, R.E., Patterson, C. and Berg, K.L.	Nucleotide sequences of moraxella catarrhalis genome	Patent: WO 0078968-A	22 28-DEC-2000;	Incyte Genomics, Inc. (US)	Location/Qualifiers																
Qy	150	TCTCAGGAAATTTACCCACCGTCGAAGTGACCAACTCGCTCAGATTTAAAAATCTGGAG	209																		
Db	28889	TCACAGCGTTAGATTCAGATCGACGAGCAGCGAGCTTGCCTTACTGTTACGCTCAGCGC	28830																		
Qy	210	CGATGTCCTTTTGTCCCGAGGTTCTCAGTGAAGACGATCTCTTCTGATCTTGGGAAA	269																		
Db	28829	CGTTAGCGGCACCGATGTACTTGTGAAGAGCGAACCATTTGGCCCATCTTTGGTCAAG	28770																		
Qy	270	AACAATGTACACAGGCATTAATCTCAGCATCTGTGGCTTGGCAATGCTTAATGTTTGA	329																		
Db	28769	ATAATATCGATTAAGGATTAATCTCTACACAGGTGGCTATCTGTTGGTATTTTATGGA	28710																		
Qy	330	TGAGCGTATATTATAGATTTTGGAGCGTCATCGCTTGGGAGCTGTTCTTCTGGAATCTTT	389																		
Db	28709	TGATTTGTTTTATCGTGCAATTTGGTGATTTGCAATATTTGCTTTAGCCCTTTAATGTCA	28650																		
Qy	390	TGCTTATCTGGGACGCTCTACAGTATTTGGATGGGCCACTCACCTTGTGAGGACTCGCTG	449																		
Db	28649	TTATCATGGTTGCCATTAATGTCCATTTTAGGATCATCGCTCACCCCTACTGTTATGCTG	28590																		
Qy	450	GGATTGTTCTTGCTATGGGAGTGGCGTAGATGCAATGTTCTTGTATTCGAAAGATCC	509																		
Db	28589	GGATTGTTGACCAATTTGGTATGGCGTTGATGCCAATGTCTAATTTTGGAGCGTATTC	28530																		
Qy	510	GAGAGGAATTTTATTGTCTCAAGTCTTAAAAATCTGTAGAAAAGGATATACCAAGG	569																		
Db	28529	GTGAAGAGCTGCCAATGGTACTCGTGCAAAATCGGCCAATTTGGCGAGTTTGTATCGTG	28470																		
Qy	570	CTTTTGGAGCCATTTTGTATTTAACTTGACTCAGATATTTGGCCCTCAGCACTTCTTTTCT	629																		
Db	28469	CATTTAGCTCAATTTTGTATGCCAATATCACCGCTATTAAATGCAATTTATTTTGT	28410																		
Qy	630	TCCTAGATACAGGCGCTATTAAAGGGTTTGTTCGATATGCAATTTTAGAATTTCTCTT	689																		
Db	28409	CCATCGGAACAGGACCAATCAAGGTTTGGCAATCGCTTGAATTTGGTATTATCAGTT	28350																		
Qy	690	CAATGTTTACGGCTCTTTTCATGACT	715																		
Db	28349	CTTTGTTTACTGCGATTTTACTCACT	28324																		
RESULT 20	AE013703	LOCUS	AE013703	DEFINITION	Yersinia pestis KIM section 103 of 415 of the complete genome.	ACCESSION	AE013703	VERSION	AE013703.1	KEYWORDS	Yersinia pestis KIM. Yersinia pestis KIM Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Yersinia. 1 (bases 1 to 10963)	ORGANISM	Yersinia pestis KIM. Yersinia pestis KIM Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Yersinia. 1 (bases 1 to 10963)	REFERENCE	Deng, W., Burland, V., Plunkett, G. III, Boutin, A., Mayhew, G.F., Liss, P., Perna, N.T., Rose, D.J., Mau, B., Zhou, S., Schwartz, D.C., Hersteth, J.D., Lindler, L.E., Brubaker, R.R						

Genome Sequence of *Yersinia pestis* KIM
J. Bacteriol. 184 (16), 4601-4611 (2002)
12142430
2 (bases 1 to 10963)
Deng, W., Burland, V., Plunkett, G. III, Boutin, A., Mayhew, G. F.,
Lise, P., Perna, N. T., Rose, D. J., Mau, B., Zhou, S., Schwartz, D. C.,
Fetherston, J. D., Lindler, L. E., Brubaker, R. E., Plana, G. V.,
Straley, S. C., McDonough, K. A., Nilles, M. L., Matson, J. S.,
Blattner, F. R. and Perry, R. D.
Direct Submission
Submitted (21-FEB-2002) Genetics, University of Wisconsin, 445
Henry Mall, Madison, WI 53706, USA
Location/Qualifiers
1. .10963
/organism="Yersinia pestis KIM"
/strain="KIM"
/db_xref="taxon:187410"
119. .1189
/gene="queA"
/note="y0990"
119. .1189
/gene="queA"
/function="enzyme; aminoacyl tRNA synthetases, tRNA
modification"
/note="synthesis of queuine in tRNA; residues 1 to 354 of
356 are 83.33 pct identical to residues 1 to 354 of 356
from E. coli K12 : B0405; residues 1 to 353 of 356 are
83.56 pct identical to residues 1 to 353 of 354 from
Genpept : >gb|AA119358.1| (AE008714)
S-adenosylmethionine-tRNA ribosyltransferase-isomerase
[Salmonella typhimurium LT2]"
/codon_start=1
/translate table=11
/product="probably S-adenosylmethionine:tRNA
ribosyltransferase-isomerase"
/protein_id="AAM84571.1"
/db_xref="GI:219577733"
/tranlation="MNRVADFSPLEALIAHYQPQSGRCRLSDGPTGLTHGIFP
DLIDKLAPDGLVFNRTVIFPARLFGKASGGKLEVLDFRVLDHRLAHVAKSKAPK
PGRELLLDGESIRATMLARHDTLELRDDEDDVETILNAGHMPLPPVIDRDEDA
DRELYVSYQRPAAVATAGLHDFEPLAALQERGIEMAFVILHVAGTGPQVRVD
TIEDHIMHSEAPVEQVVDVALACKARKGRVAVGTTSVRSLESAAKAAENGLIAPF
FIVTGYFYGHYQVVDALVNTFHLPESTLMLVSAFAGYKNTMAYQQAAVEQYRF
FSYGDAMPISRNPRAQPEKVP"
1282. .2406
/gene="tgt"
/note="y0991"
1282. .2406
/gene="tgt"
/function="enzyme; aminoacyl tRNA synthetases, tRNA
modification"
/note="residues 1 to 374 of 374 are 89.57 pct identical to
residues 1 to 374 of 375 from E. coli K12 : B0406;
residues 1 to 374 of 374 are 89.83 pct identical to
residues 1 to 374 of 375 from Genpept : >gb|AA119359.1|
(AE008714) tRNA-guanine transglycosylase [Salmonella
typhimurium LT2]"
/codon_start=1
/translate table=11
/product="tRNA-guanine transglycosylase"
/protein_id="AAM84572.1"
/db_xref="GI:219577734"
/tranlation="MKYELQKTGDRARGLYPERGVETPAMPVGTCTYTKVGMTPP
EVKETAQLLNGNTFLHWRPQGEIMKLHGDHDFQWKGPIITDSGGGFVFSLGAWPR
KILKEGVHPKPNQFSDLSPEKSMIEIYDLGSDIVMIFDECTPYADNAYAKRSGV
MSLRBAWQRINGFDLNNKALFGIIQQGVYEDLRDVSVKGLVDIGFDGAVYAKRGLV
EPKQDMHRLTLEHVCPQIPDKPRYLQKGPEDLVKGVRRGIDMFQCFVMPTRNARNGH
LFTVDGVKIRNAKHKKDDTATLDEHCDQYTCRHSYRAYLHLDRCNEILGARLNTIAP
LRYQLRMAGLQRAIEEGKLEHVEFDYGRICKVPPLNV"
2452. .2853
/gene="y0992"
2452. .2853
/gene="y0992"


```
Db 4526 CCTTCTCAAGTATCGTCGATCGGAACAATAACGACGCTTATCACAGCCATTATCTTTATG 4585
Qy 630 TCCTAGATACAGGCGCTATTAAAGGGTTTGGCTTTGACATGATTTAGGAATTTTCTCTT 689
Db 4586 CTGTGGTACTGTTTCGATTAAAGGCTTTGCTATTACACGCAATTGTTGTTAAAGT 4645
Qy 690 CAATGTTTACGGCTCTTTTCATGACTAAATTTTCTTCATGCTGTGGATGAATAAGACCC 749
Db 4646 CAATGTTTACCGCAATAGTCGATCTCGTGCCATCGTCAACCTGCTTACGCGCGCAAGC 4705
Qy 750 AACATACACAGTGCATATGATGATAGTTCGCGGATAAAGCATGATTTCTTGAGAG 809
Db 4706 GCATTAAACAGCTGTCTAT--TTGAGGAGTACGTTGTGGCAGAGTTATCTGTTGAAC 4763
Qy 810 GATGCAAAAAAATTTGGCTGTTCTGGAAGTGTTTTCTTTAGTTGCGTGTCTCTCG 869
Db 4764 AACTAAACTATGGCGGTAAGCTATGACT--TTATGCGGTGGGATTACGTTGCTTCG 4820
Qy 870 GGTTTGGAGCTGGAATTCGGTTTGGGAATGGAATTTAAAGGAGGTATGCTTTTACCT 929
Db 4821 CGGTTTCACTGTTGTTGTAGTCGCATCCATTTGGTGATGTCGACTAAGGGAATCAACT 4880
Qy 930 TTAATCCAAAGAGCATGCGATCAGCATGTTGCTCAATCGTGGCAGAAAGTTGTGCATA 989
Db 4881 GGGGTCTGGATTTACCGGTGGTACGGTGATGAAATTAATCTGGAAACCCCTCGCGAT- 4939
Qy 990 AACTACAGGAAGCTGTTCTTCTCTAGAGACTTCCGTATTCAACATTTGGATCTTCAG 1049
Db 4940 -----CTTGACCAACTCGTGACAGCTGTCGAAACGGGGTTTGAATCGCGG 4987
Qy 1050 AAAAGATCAAAATCTATTATTAGTATAAGCTTTAAGCTATATACTAAGCAGATACGAGCT 1109
Db 4988 ATACTGCAAAACTTTGGT-----AGCTCTCGTATGTCATGTGTACGTATGCCACC 5037
Qy 1110 CTCTCCTAAATTAACGATCATGAGCTGGCTTATTGTGGGATTTGTTGTGAGAACAGGC 1169
Db 5038 TGCAACCGGCACTGCGCGCCAGGAGCTCG-GTAACAAAATCATCTCGGTTATCAATGAAT 5096
Qy 1170 CTAGATTTCTCTACGGAACCTTAACGAAAGCGCAAAATTTTGGTCAAGGTAAAGCAGCA 1229
Db 5097 CTGTTGATAGATGATCAGTAAGCGATATGAAATTCGTTGGGCTAGTGTGGCAGTG 5156
Qy 1230 AACTATCGAAGAAATGCGTTATACGCGCAGCATCGGCTTTTAGGAGCTTTGGCAATCA 1289
Db 5157 ATTTAGC-----CCAAAGCGGTGCTCTGGCGCTTTTGTCTCGCTTACTGTCTA 5204
Qy 1290 TCTTGCTCTATGTAGTTTGGCTTTGAATGGCAATATGCTTTAGTCGCGTATGCGCTT 1349
Db 5205 TTCTGGTTTACGTGGGTTTCCGTTTGAATGGCGTCTGGCATTGGGGCGCGTTATCTCAT 5264
Qy 1350 TAATTCATGACCTTTTGGCTACCTGTGAGTCTTCTTTTATAGCACATTTCTTTTGAAGA 1409
Db 5265 TAGCCCATGAGTGTGCTATTACCATGGGGATCTTG-----TCGTTGTTC 5309
Qy 1410 AAATTCAAATAGATTTGAAGCAATTTGGTGTCTTTAATGACTGTATTTGGGGTATTCATTAA 1469
Db 5310 ATATTGAGATTTGACCTGACGATCATTTGCTTATTGATGTCGTTATTGTTACTCTCTGA 5369
Qy 1470 ACAATACTTTGATCATTTTGTGATGTTATGTTGAAGATGCGCAAGCAAGCACTGTTTACC 1529
Db 5370 ACGATAGTATCGTGTGATCAGACCGCTATTTCGTGAGAACTTCGCGCAAAATTCGTCGTGAA 5429
Qy 1530 CTATGCATGTTTATGTTAATGATGCTTCAAAAGACGCTTCAGCGCGCATGTAATGACAA 1589
Db 5430 GCGCTTAGCAAAATCAAGATGATCTCTGACACAAACCTTAAGTCGTACGATTATGACTT 5489
Qy 1590 CAGCTACAACCTCTATCAGTTTTTGTAAATGCTTTTCTTTTATAGCGCGCTCTCTGCTTTA 1649
Db 5490 CTGCGACGACTCTGATGTTGTTGTGATGCTCTTTTATCTTCGGTGTGGATGCTGCAAG 5549
Qy 1650 ATTTTGCATTTATTATGACCAATAGGAGTCTTCTAGGAACCTTTATCGTCTCTTTATATTG 1709
```

```
Db 5550 GCTTCTCATTAACGATCGTATCGGTGTTACTACTCGGTACGAGCTTTCATCTATCTACGTGG 5609
Qy 1710 CACCACCTCTGTTGTTG 1726
Db 5610 CGTCCGCGCTGGCGCTG 5626

RESULT 21
AJ414155/c
LOCUS
DEFINITION
  Yersinia pestis strain CO92 complete genome; segment 15/20.
ACCESSION
  AJ414155 AL590842
VERSION
  AJ414155.1 GI:15980975
KEYWORDS
  Yersinia pestis.
SOURCE
  Yersinia pestis.
ORGANISM
  Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
  Yersinia.
REFERENCE
  1 (bases 1 to 205050)
  Parkhill, J., Wren, B.W., Thomson, N.R., Titball, R.W., Holden, M.T.G.,
  Prentice, M.B., Sebaihia, M., James, K.D., Churcher, C., Mungall, K.L.,
  Baker, S., Basham, D., Bentley, S.D., Brooks, K., Cerdano-Tarraga, A.M.,
  Chillingworth, T., Cronin, A., Davies, R.M., Davis, P., Dougan, G.,
  Peltwell, T., Hamlin, N., Holtroyd, S., Jagels, K., Leather, S.,
  Karlyshev, A.V., Moule, S., Oyston, P.C.F., Quail, M., Rutherford, K.,
  Simmonds, M., Skelton, J., Stevens, K., Whitehead, S. and Barrall, B.G.
  Genome sequence of Yersinia pestis, the causative agent of plague
  Nature 413 (6855), 523-527 (2001)
  21470413
  MEDLINE
  11598360
REFERENCE
  2 (bases 1 to 205050)
  Parkhill, J.
  Direct Submission
  Submitted (04-OCT-2001) Submitted on behalf of the Yersinia
  sequencing team, Sanger Centre, Wellcome Trust Genome Campus,
  Hinxton, Cambridge CB10 1SA E-mail: parkhille@sanger.ac.uk
  Notes:
  Details of Y. pestis sequencing at the Sanger Centre are available
  on the World Wide Web.
  (URL, http://www.sanger.ac.uk/Projects/Y_pestis/).

FEATURES
  source
    location/Qualifiers
    1..205050
    /organism="Yersinia pestis"
    /strain="CO92"
    /db_xref="taxon:632"
    /note="biovar: Orientalis"
  gene
    69..1197
    /gene="YPO3021"
    /pseudo
  CDS
    69..1197
    /gene="YPO3021"
    /note="Similar (after the first 34 N-terminal aa) to the
    previously sequenced Yersinia pestis outer membrane
    protein YopM or Yop48 SW:YOPM_YERPE (P17778) (367 aa)
    fasta scores: E(): 3.9e-25, 37.5% id in 339 aa, and to
    Yersinia enterocolitica YopM protein TR:P74988
    (EMBL:AF102990) (367 aa) fasta scores: E(): 8.2e-26, 37.2%
    id in 339 aa. There is a possible frameshift following
    codon 188. The frameshift occurs at a homopolymeric tract
    of 8C. The sequence has been checked and is believed to be
    correct"
    /pseudo
    /codon_start=1
    /transl_table=11
    /product="putative outer membrane protein (pseudogene)"
    258..326
    /gene="YPO3021"
    /note="Pfam match to entry PF00560 LRR, Leucine Rich
    Repeat, score 14.40, E-value 2.8"
  misc_feature
    /pseudo
    378..443
    /gene="YPO3021"
    /note="Pfam match to entry PF00560 LRR, Leucine Rich
```

```

misc_feature
Repeat, score 16.20, E-value 0.79"
/pseudo
447. .512
/gene="YPO3021"
/notes="Pfam match to entry PF00560 LRR, Leucine Rich
Repeat, score 16.30, E-value 0.71"
/pseudo
516. .581
/gene="YPO3021"
/notes="Pfam match to entry PF00560 LRR, Leucine Rich
Repeat, score 17.80, E-value 0.26"
/pseudo
1403. .2104
/gene="YPO3023"
1403. .2104
/gene="YPO3023"
/notes="Similar to Escherichia coli hypothetical 24.1 kDa
protein in gltF-nant intergenic region YhcJ SW:YHCJ_ECOLI
(P45426) (229 aa) fasta scores: E(): 0, 59.7% id in 226
aa, and to Pasteurella multocida conserved hypothetical
protein YhcJ TR:Q9L6B4 (EMBL:AF237924) (228 aa) fasta
scores: E(): 0, 58.2% id in 225 aa. It is also similar to
the N-terminal region of Streptomyces thermoviolaceus
beta-N-acetylglucosaminidase precursor Nagd TR:082840
(EMBL:AB008771) (632 aa) fasta scores: E(): 0.86, 25.1% id
in 207 aa"
/codon start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="CAC92265.1"
/db_xref="GI:15980976"
/db_xref="SPTREMBL:08ZCG8"
/translacion="MSNLSNLRHKLQNGLIASCQVPGSAMDPETVVAAMACALAG
AVGLRLEGISNIQAVRRATDAPITGIKRLDPSEVRIIPWEDIDALSGAGDIATF
DYTCRRPVSVAIDYORARATGCTTMADASNIDGDLAHLGIDFTGLSGYQATV
PREPDLAVTOLAAGCRVIAEGRHSPALAAAISGAAVAVGSATRIEHIICMF
CDAIKOCETHEKLEY"
2167. .3060
/gene="YPO3024"
/notes="nana; np1"
2167. .3060
/gene="YPO3024"
/EC_number="4.1.3.3"
/notes="Similar to Escherichia coli N-acetylneuraminate
lyase subunit Nana or Npl SW:NPL_ECOLI (P06995) (296 aa)
fasta scores: E(): 1.7e-16, 29.1% id in 265 aa, and to
Vibrio cholerae N-acetylneuraminate lyase, putative VC1776
TR:Q9KR67 (EMBL:AE004255) (298 aa) fasta scores: E(): 0,
75.1% id in 293 aa"
/codon start=1
/transl_table=1
/product="probable N-acetylneuraminate lyase"
/protein_id="CAC92266.1"
/db_xref="GI:15980977"
/db_xref="SPTREMBL:08ZCG7"
/translacion="MKKLTGLIAAHPPTPEDEGVNYPVIDIAEHLINDGKGVYVC
GTTGEGHLSVDERKKIAERWNAOQKSTITLATALSKDANVDSRAHETADITAT
SAIGPCFFKPGKNDLIAYCOALAAAPSKFYHSGMSGVNIDMEQFLIKESKIP
NLSGIKENNADLYEFORCLAVSGKFDIPGVGVEHLPGLAVGAIAGVSTVYAPL
FKKTIADFNAGDOVAORGMHDVITLIRIVVERGVGAAGAAQMLGIDAGNRLRLR
ALTKEQKQIVVNMRAITIQ"
2236. .3300
/gene="YPO3024"
/notes="Pfam match to entry PF00701 DHDPS,
Dihydrodipicolinate synthetase family, score 87.30,
E-value 3.1e-22"
2284. .2337
/gene="YPO3024"
/notes="P800665 Dihydrodipicolinate synthetase signature
1."
complement (3158. .4057)
/gene="YPO3025"
complement (3158. .4057)

```

```

gene
CDS
/gene="YPO3025"
/notes="Similar to Escherichia coli hypothetical protein
B2431 TR:P76536 (EMBL:AE000330) (308 aa) fasta scores:
E(): 0, 66.7% id in 300 aa, and to Pasteurella multocida
hypothetical protein PM0455 TR:AAK02539 (EMBL:AB006081)
(296 aa) fasta scores: E(): 0, 65.0% id in 297 aa"
/codon start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="CAC92267.1"
/db_xref="GI:15980978"
/db_xref="SPTREMBL:08ZCG6"
/translacion="MTOYOSGILHCEFAIPMEAKVOGELDPAIRLQKPKCSLOEL
OOQPFDEHGAIVARGSNVMHDSMGQAKELKPVPGKGLAPATQDILIHQSLR
QDINTTLQAAVAFGSNAIABEETHGRVWEERPTGIDSTENPQDKPREAVATA
DGEEDAGSYLVQVYEHNLNWKRIPENEDEKIGRTKLDSQELPSQDPPTSHVR
VDLKNKGKGLKILRQSLPYGLASGHGGLFYACARLHNIEQOLISMWDIDGKHQOL
LRFSPVTSYVFAPSLTALLSL"
complement (4235. .4846)
/gene="YPO3026"
/notes="YfeY"
complement (4235. .4846)
/gene="YPO3026"
/notes="Similar to Escherichia coli hypothetical 20.9 kDa
protein in ucpa-amlia intergenic region precursor YfeY
SW:YFEY_ECOLI (P76537) (191 aa) fasta scores: E(): 0,
56.5% id in 193 aa, and to Aeromonas hydrophila blr
regulon signal sensor BlrB TR:Q9K96 (EMBL:AJ276632) (473
aa) fasta scores: E(): 0.65, 28.8% id in 163 aa"
/codon start=1
/transl_table=1
/product="putative lipoprotein"
/protein_id="CAC92268.1"
/db_xref="GI:15980979"
/db_xref="SPTREMBL:08ZCG5"
/translacion="NNIRPLRSGVLLLSAPLLIAGCSTVSGFSMSHSPINWFGS
RSMOYBOGVGIGTASTLTLESDINAGOGDYRLASGMATNDGKLVSPFOALKEQIK
LVISQPRQVQRIDVMDKIPISQMGVITGTFPSDLYKKARACRKGKGDVNOVECV
APESKRVSTLEFNGHMGPELMSDIALQSKVSKIMPAKSE"
complement (4775. .4807)
/gene="YPO3026"
/notes="P800013 Prokaryotic membrane lipoprotein lipid
attachment site."
5205. .5618
/gene="YPO3027"
/notes="YgiW"
5205. .5618
/gene="YPO3027"
/notes="Similar to Escherichia coli protein YgiW precursor
SW:YGIW_ECOLI (P52083) (130 aa) fasta scores: E(): 1.3e-24,
52.9% id in 136 aa, and to Bacteriophage 933W,
and Bacteriophage VT2-8a hypothetical 12.4 kDa protein
10136 TR:Q9XJM3 (EMBL:AF125520) (114 aa) fasta scores:
E(): 7.5e-16, 51.0% id in 96 aa"
/codon start=1
/transl_table=1
/product="putative exported protein"
/protein_id="CAC92269.1"
/db_xref="GI:15980980"
/db_xref="SPTREMBL:08ZCG4"
/translacion="MKKLTIAATLIMLCSTPYLAQASNTTQSGFDPATTOPTOATO
GGFSPASASTTVDKYKTRMDRPAATLQGYIEARLGHDTYTRDSTGTTTVEIDKRW
RGQTTTPQDKVIEKVKDMNSVEIDVKNITKIX"

```

Query Match 6.2%; Score 115; DB 1; Length 205050;
 Best local Similarity 46.2%; Pred. No. 26-15;
 Matches 728; Conservative 0; Mismatches 790; Indels 59; Gaps 8;

QY 153 CAGGGAATTATCCACCGTGAAGTAGCAAACTCCCTCAGATTTPAAATCTGAGCGCA 212
 DB 175914 CCGGCATTGATTAACCGAGCGAAGCCGCGCACTTCCGTACTACGCGTGGTGCGC 175855
 QY 213 TGTCTTTTGTCCGAGGTTCTCAGTGAAGAGACGANTCTTCTGTGANTCTTGGAAGAAAAAC 272

goal is to make the genome sequence more useful. Comments to the authors are appreciated.

```

FEATURES
    source
        location/qualifiers
            1..10899
            /organism="Rickettsia conorii"
            /strain="Malish 7"
            /db_xref="taxon:781"
            complement(103..131)
            /note="REP01, repeated element"
            154..1098
            /gene="RC0885"
            154..1098
            /gene="RC0885"
            /codon_start=1
            /transl_table=11
            /product="unknown"
            /protein_id="AA03423.1"
            /db_xref="GI:15619991"
            /translation="MTCGLPSRLILAMTISIDSSOYOISRTNMLIKLKHIPDVNDY
            DVFLFDIMGVIITGGCTTGGVONINKITIERKAVYTYTNAPRNIFSLHQTIKSGLNA
            EPGMIITSSGKIIVEMILESKEKPGIEKPAIVHGLHNDIINSIQPTIDINKANIF
            LMTIYRDENMLDNEFENELPKIVORKNVINCANPDLGINOHGYRKYRCSGYAEKIK
            QLGKVIYSGKPYEIEYSKIKECHNPKRMIMIGDVTYTDILAA NRIGIDISALVLT
            GNSREYHIEFDNIDEXKLSMKASVKOSITPSFVVDLS"
            complement(161..194)
            /note="REP01, repeated element"
            repeat_region
                gene
                    1253..1336
                    /gene="RCRNA24"
                    1253..1336
                    /gene="RCRNA24"
                    /product="tRNA-Leu"
                    1604..2574
                    /gene="mctc"
                    /note="RC0886"
                    1604..2574
                    /gene="mctc"
                    /codon_start=1
                    /transl_table=11
                    /product="magnesium transporter"
                    /protein_id="AA03424.1"
                    /db_xref="GI:15619992"
                    /translation="MPNFNIFKGIILPDHDOFAEIFDQINDLIEDHNYDAASRLAKL
                    HYADLADPRLDNLNKKYKTIITPILODEIKRETVISNAYSKPLIETLIGIKSAEILIN
                    KLVIEDAIVAVITDLDIDDKLILSNLTKKKEQITVCTTPEPTVGRVIRERDINQOE
                    GWTVESLEINFLIKVKNDFYAVIATDNKLRPIGISMLIKGKONELIKLAKKDFRL
                    ADAFTDLNLSLPIFERQYALTIYVYVNSGKLVSVSIDNIYIIEQAGDILISLBSG
                    HTODTFENVYTYVARPPMLFVNLITACMTSLIINHNDIATLVTLLAATMPDILSVMG
                    GNAGTORVTVRALANKDIIHNNVKNVILKEITAVSAGFNGFVLAITIGAGSFFVMLDL
                    NLSVIRAIIVILNPLVAGLFGSAPILIHFDIDPATGSGVPLITITDNLGFLTFISL
                    AHIPLV"
                repeat_region
                    gene
                        complement(2978..3154)
                        /gene="R0887"
                        complement(2978..3154)
                        /gene="RC0887"
                        /codon_start=1
                        /transl_table=11
                        /product="unknown"
                        /protein_id="AA03425.1"
                        /db_xref="GI:15619993"
                        /translation="MPKESNINNHQNIQRDNNISYENNITTPDENQDTSGLCKICGA
                        IATVCCNVYVPIFE"
                        complement(3316..3482)
                        /note="REP06, A and T rich repeated element"
                        3823..4275
                        /gene="RC0888"
                        3823..4275
                        /gene="RC0888"
                        /codon_start=1
                        /transl_table=11
                        /product="similarity to (p)ppGpp 3-pyrophosphohydrolase"
                        /protein_id="AA03426.1"
                        /db_xref="GI:15619994"
                        /translation="MRQSGNPIYSHPIEIKIMLAEPVAEAPKLVNIMIQVQALILH

```

```

gene
    complement(4315..4656)
    /gene="RC0889"
    complement(4315..4656)
    /gene="RC0889"
    /codon_start=1
    /transl_table=11
    /product="similarity to proline/betaine transporter"
    /protein_id="AA03427.1"
    /db_xref="GI:15619995"
    /translation="MVALMSPILNNSITTELLFOLPIVFAPIFPAGAVFYARF
    PVLRKPTGSPFIALSRALMFAVSSFRITLYLDYRNHMLLELIPILIGYFGVNHFI
    KLEEEASVY"
    complement(4978..5490)
    /gene="RC0890"
    complement(4978..5490)
    /gene="RC0890"
    /codon_start=1
    /transl_table=11
    /product="similarity to proline/betaine transporter"
    /protein_id="AA03428.1"
    /db_xref="GI:15619996"
    /translation="MAVILNELFPEDADAPASLYSALAYCATVYFRPICALIFGMIG
    DTIGKRAVYITTFPLMSICVILMANYKNLGNMYGYPCNCVGYRSGDVLVYRSSR
    CRALSNRDNKNSSTILGGFACCFNFCNRVALAVGTLVTSFGDLTYCIIYMSYSY
    YSWCCCKDKS"
    6025..6096
    /gene="RCRNA25"
    6025..6096
    /gene="RCRNA25"
    /product="tRNA-Gly"
    6122..6146
    /note="REP02, RS3-like repeat"
    complement(6239..6397)
    /gene="RC0891"
    complement(6239..6397)
    /gene="RC0891"
    /codon_start=1
    /transl_table=11
    /product="unknown"
    /protein_id="AA03430.1"
    /db_xref="GI:15619997"
    /translation="MAIMFTVITTSINSELDIIRKAIENYIEKADLLIALSCIEIKR
    KSNFTIRY"
    6506..6647
    /note="REP05, Rickettsia palindromic element 1"
    6605..6844
    /gene="RC0892"
    6605..6844
    /gene="RC0892"
    /codon_start=1
    /transl_table=11
    /product="unknown"
    /protein_id="AA03431.1"
    /db_xref="GI:15619998"
    /translation="MNTYANISKVSGSISGEPARERKLIHQRRSLKFDVYKLEVSQVY
    KLLERGYTRLLQDLGNVTCQRIKRIIRQIITTL"
    6796..7221
    /gene="RC0893"
    6796..7221
    /gene="RC0893"
    /codon_start=1
    /transl_table=11
    /product="unknown"
    /protein_id="AA03431.1"
    /db_xref="GI:15619999"
    /translation="MSANTQDNQANNNDTIEIOETENVYPAETNSIQSLTSLIPWLI
    FAVYFLLRPQETRRERERKQVSEVKGEVLTNSGIYITKVSSENDNNIEITIAK
    DVRIYVLSAIIIDITSRTKEVAVVKKENKKRVSGAKAS"
    7205..8761
    /gene="secD"
    /note="RC0894"

```

CDS	7205..8761 /gene="secD" /codon_start=1 /transl_table=11 /product="protein-export membrane protein secD" /protein_id="AAL03432.1" /db_xref="GI:15620000" /translation="MQKLPRWKIFLSIICTFVAVICALPNFMQVNSFLPHSDVNLGL DLRGHLLLDVDEDTYLDNSMENLADTLRNFREDKIGYNLLVLRONSIOLEVRSE KLRLKKIINKIDPEIIYAENVENKILSYSESLNDLLNKVVDQSIIVIMRVDSGT SEPTLQGGQKHILLQVGENPSYLNILGKTAKLTFLHVDENATIEAVKGVHPVG KMLVKGASHGEYVVVKKVVLGGDPLTTASAPDQNSOAVAFSNNLGSKIFGE ITKNICKRLAIVLDNKLSPATINGAIMGSGIITGNFTVESNELALLIRAGSLPA PLKIIERSIPGNLGAUSIESGKAGLGIYVAFICFVFWWSYGVGLFANIALSLALDY
Query Match	6.0%; Score 112.6; DB 1; Length 10899;
Best Local Similarity	51.7%; Pred. No. 9.4e-15;
Matches 256; Conservative	0; Mismatches 239; Indels 0; Gaps 0;
Qy	284 GGCATTATCTCAGCATGCTGTGGCTTGGCAATGCTTATTTGTTGATGAGCGGTATATTAT 343
Db	8270 GGC AAAAGGCGAGCATTTATAGGATTTTATAGCAGTGTGTATATTATGTTGGTCTTAC 8329
Qy	344 AGATTGGAGCGCTCATCGCTCGGGAGCTGTCTTCTGAATCTTTGCTTATCTGGCA 403
Db	8330 GGTGTGCTGTTATTTGGCAATATAGCCTTAAGCCTTGCAATGTATATATTTAGCT 8389
Qy	404 GCTCTACAGTATTTGGATGCGCCACTCACCTTGTGAGGACTCGCTGGGATTTGTTCTGCT 463
Db	8390 TTACTTTTCAAGTACTTTTAACTCTGCGGGAATGTCAGGAATAATACTCACT 8449
Qy	464 ATGGGATGGCGTAGATGCAAAATGTTCTGTATTCGAAAGAAATCCGAGAGAAATTTT 523
Db	8450 ATGGGTATGGCTGTGCTAAATGTTATGTTTACGAAAGAAATTAAGAGGAACTGCAT 8509
Qy	524 TTGTCTCAAGTCTTAAATAATCTGTAGAAAAGATATACCAAGGCTTTTGGAGCAATT 583
Db	8510 AAAGAGTGTCTACTTTTATGCTATTAGAACCGGTTTGAATCTGCGTTTGGCCACTATC 8569
Qy	584 TTGATTTCAACTTGACTACAGTATTTGGCTCAGCACTTCTTTTCTTCTAGATACAGGG 643
Db	8570 CTTGATGCTAACCTAACCTACTTAACTGTTGCTTTTACTTTATATATTTGGAGTAGT 8629
Qy	644 CCTATTAAGGGTTGCTTTGACATGATTTTAGGAATTTTCTCTCAATGTTTACGGCT 703
Db	8630 GCAATAAAGGTTTGGCGTGGCAATTAACGATCGGATTTATATCTTCAATGTTTTCGCA 8689
Qy	704 CTTTTCATGACTAAATTTTCTTCTCATGCTGCGATGATAGACCCACATACACAGTTG 763
Db	8690 ATTATCATTAATGTTTAAATGATATTTGGGTACAATATTTTAAACCTTAAAAAATTA 8749
Qy	764 CATATGATGAATAAG 778
Db	8750 GGGCTTGTTGAAG 8764
RESULT 23	
AE001166	13990 bp DNA linear BCT 15-DEC-1997
LOCUS	AE001166
DEFINITION	Borrelia burgdorferi (section 52 of 70) of the complete genome.
ACCESSION	AE001166 AE000783
VERSION	AE001166.1 GI:2688571
KEYWORDS	
SOURCE	Borrelia burgdorferi.
ORGANISM	Borrelia burgdorferi
REFERENCE	Bacteria; Spirochaetales; Spirochaetaceae; Borrelia; Borrelia burgdorferi group.
AUTHORS	1 (bases 1 to 13990) Fraser, C.M., Casjens, S., Huang, W.M., Sutton, G.G., Clayton, R., Lathigra, R., White, O., Ketchum, K.A., Dodson, R., Hickey, E.K., Winn, M., Dougherty, B., Tomb, J.F., Fleischmann, R.D., Peterson, J., Kerlavage, A.R., Quackenbush, J., Salzberg, S., van Vugt, R., Palmer, N., Adams, M.D., Gocayne, J., Artach, P., Bowman, C., Garland, S., Fujii, C., Cotton, M.D., Horst, K., Roberts, K., Hatch, B., Smith, H.O. and Venter, J.C.

TITLE	Venter, J.C. et al.
JOURNAL	Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi
MEDLINE	Nature 390 (6660), 580-586 (1997)
PUBMED	98065943
REFERENCE	2 (bases 1 to 13990)
AUTHORS	Fraser, C.M., Casjens, S., Huang, W.M., Sutton, G.G., Clayton, R.A., Lathigra, R., White, O., Ketchum, K.A., Dodson, R., Hickey, E.K., Winn, M., Dougherty, B., Tomb, J.F., Fleischmann, R.D., Richardson, D., Peterson, J., Kerlavage, A.R., Quackenbush, J., Salzberg, S., Hanson, M., van Vugt, R., Palmer, N., Adams, M.D., Gocayne, J.D., Weidman, J., Uterback, T., Watthey, L., McDonald, L., Artach, P., Bowman, C., Garland, S., Fujii, C., Cotton, M.D., Horst, K., Roberts, K., Hatch, B., Smith, H.O. and Venter, J.C.
TITLE	Direct Submission
JOURNAL	Submitted (12-DEC-1997) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA
FEATURES	Location/Qualifiers
source	1..13990 /organism="Borrelia burgdorferi" /db_xref="taxon:139" complement(70..909) /gene="BB0643" complement(70..909) /genes="BB0643" /notes="similar to GB:L43967 PID:1046160 SP:Q49435 percent identity: 38.94; identified by sequence similarity; putative" /codon_start=1 /transl_table=11 /product="conserved hypothetical protein" /protein_id="AAC67000.1" /db_xref="GI:2688581" translation="MANKINWFGHKRALDLIKNNLQKANIIVLEILDARAPSSKNP LTKEITKNQAKIILLKHSVDAQINEIILKWKYFENLGNVTIISNIYKGGKQIIDI KKLAIYKKIKNYKEKIKVIGVNNVKSIIINLLSGKSAKVANKPYTKNQIIVKI NEINFDMPGLMHNLDQSIQAKLAILDIMEIIVDNTDLALYLEIMDQNNKIL LKKEIYVHKNSLDILQNFARKARKLIKGNELNELEKASIKILKEFREGKFGKILDKNY NAT"
gene	1064..1762 /genes="BB0644"
CDS	1064..1762 /genes="BB0644" /notes="similar to GP:1573101 percent identity: 36.77; identified by sequence similarity; putative" /codon_start=1 /transl_table=11 /product="conserved hypothetical protein" /protein_id="AAC66999.1" /db_xref="GI:2688580" translation="MLIIYIKIKGLIVSQALENELHSSFINSKWLAARKIGGAI RANGVNDISQIKLEVDLPILGIKKNNYCNDFITPTMKEDISLCEGVDIILADATP RNPDGLVLDFFENIKKYPKQCLMADISLDEAINADKGFDFITGLTYTKNTN GLNIADNDFNLRLTLLNSLNLSTLIVEKIDTFLKAKCFEMGVDLVVVGATIRPAE ITKKFVEKINQIKK" 1808..3352 /gene="BB0645" 1808..3352 /genes="BB0645" /notes="similar to GB:X80415 PID:1072418 PID:515384 percent identity: 46.03; identified by sequence similarity; putative" /codon_start=1 /transl_table=11 /product="PTS system, glucose-specific IIBC component (ptsG)" /protein_id="AAC66998.1" /db_xref="GI:2688579" translation="MLKGFQAKQCFGRFWMPLPIAILPAAGLLIGIGSLSNPETVRY SPNIFLQSVFKIMASGSIIFSNLAPIFISGIAVLAKSDKGTSGIAAFIYGLVNN ATIGVLIDVSGRAESFSSGAVLKITETGFCGVVVGILTYLHRSFNKVDLPK VLGFFSGSRFPVIIVSFSIFLAVIMFLVMPFVQSGINKVGLVDSTGYTIGTYGIF

```

gene
CDS
    complement (3382..4365)
    /gene="BB0646"
    /complement (3382..4365)
    /gene="BB0646"
    /note="hypothetical protein; identified by Glimmer; putative"
    /codon_start=1
    /transl_table=1
    /product="B. burgdorferi predicted coding region BB0646"
    /protein_id="AAC67003.1"
    /db_xref="GI:2688584"
    /translation="MNKNIIIFILFLLILIVSPRIKFNKESKLLIPANIIEIDNY
    LKEELQFNLESNTKEIIMYKEKAKTNSVYVHIGFQSKNEIYPPVNIILAKALNA
    NIFTRLKRGHGNKNAFRGITTQDMRIDEMAINIGKLDGLVIGISNGSTAIW
    ALANYPNEINSAVLISPNIFPYDKRTNIVYPMGRQIAYLITGGINKFETKEKKEH
    PIKSHSSRYQVDAIIMAGLVTLNSVNFENIKLPLIITHPNHTYDPIKINFEI
    KNYGKKCIPIILLNSHALPIGNQSYKSAQNTSYFTKYVPDFINKINK"
    complement (4362..4892)
    /gene="BB0647"
    /complement (4362..4892)
    /gene="BB0647"
    /note="similar to GP:1667516 percent identity: 25.19; identified by sequence similarity; putative"
    /codon_start=1
    /transl_table=1
    /product="feric uptake regulation protein (fur)"
    /protein_id="AAC66997.1"
    /db_xref="GI:2688578"
    /translation="MNDNIDVHSALEKVGITNDPYLLKNLISELGMKASHNRNII
    YLANSKEYTAKEVYKIKELIPSLSKATVYNTLTKERNILKDKITDOKEKFEY
    LSLASTIAHKCKNCQVHPIQDDIKDLKDLGEMWETKSEIITISGHCNNCKYKD
    THNNNVPIENKEITL"
    complement (4992..6680)
    /gene="BH0648"
    /complement (4992..6680)
    /gene="BH0648"
    /note="similar to GP:887563 percent identity: 30.22; identified by sequence similarity; putative"
    /codon_start=1
    /transl_table=1
    /product="serine/threonine kinase, putative"
    /protein_id="AAC66996.1"
    /db_xref="GI:2688577"
    /translation="WFRKLKKLTYIIIIISYRSEDNLLIISNIIYLEHQIAYK
    IHMTPIYFPEILREHELMKMLPERKNTVDIYMPGTYSGSPEHYMLHDEIHDLTW
    ALKNPFKSYKXIDIFONTPIFYIYNIIEKPKKYTELARKLNFYTBGIRQSKNNKYL
    IEPKNNCOCLYEVOIKIDSPKSNVETLIYFEIKEYDNOKLKNFLYLKALLENKHDI
    KIONLESSCLTTELEIIPKNSLKEOEPINPONKRIKIDOKESLREPLIKHDOE
    IIRNSBSIYPRKNLEYMNEGFTLSDQVYMKFPGNKIKNKIKPKDKVEPEPNTSRTEK
    VSSKEKLEIKASIESFSFSNEKILGIGKYLAFNSAKSITDPFIDETISYISIKI
    KMPKSIDLCKLTKKCNPDYLLKYSALIEFETIKGTNITAKYSDDLTYSKIIIT
    KNNPKGIVGTEFLISKDNKSNFISPLNVEKHIIHTINYIKEXKNSKMWLLING
    GSYNTKIDGVINVSQTLNMLPLANNPNKIKLKNKIKNLIIFYTNIKKYENK"
    6875..8512
    /gene="BB0649"
    /complement (8512..8512)
    /note="similar to SP:P27575 GB:X54059 GB:X65139 PID:39365
    PID:39374 percent identity: 99.45; identified by sequence
    similarity; putative"
    /codon_start=1
    /transl_table=1
    /product="heat shock protein (groEL)"
    /protein_id="AAC66995.1"
    /db_xref="GI:2688576"
    /translation="MAKDIYFNEDARKSLISGVKLSNAVKVTLGPKRNVLLDKKRG
    SPVTKDGVAVAREIELNPFENKGAQLAEVAIKITNDVADGDTTATVLAVALARBG

```

```

LKNVSGINPIGKIDHVAVLAERIKROSARKITTEKEBIAQVASISANDSYIGEK
IAEMDKYGDGVLTIVESKTEPDTLIVSGEQPGRGLISPESFNKEMSVNPDPAF
ILIIYKSKISIKELDLIVEXVLGNTKPLLIITAEDESGALALVANSRGALKVCAIK
SPGFEDRRKAMLELDIAVLITGLVISEEGLTELEIYQIQGAKTIRYDKDVTIINT
GNKEQIKRSEILIKQIEDSTSEYDKELQOEALVLGVGVANVGAATEVELKREKH
RVEDALSATRAVAEEGVVPGGSLIEVAMYLDTIDSKLSTYEEKQGEIYRSLIEEP
MROIISNAGFESGIYHOIKTEKGLAGDASFQKVMNIESGIIIDPAKVTISALONAA
SIAGLLITTECAITDIEKEKNTSGGGVPMPOGMGM"
5336..8874
    /gene="BB0650"
    /complement (8874..8874)
    /note="hypothetical protein; identified by Glimmer; putative"
    /codon_start=1
    /transl_table=1
    /product="B. burgdorferi predicted coding region BB0650"
    /protein_id="AAC67002.1"
    /db_xref="GI:2688583"
    /translation="MFLPVNILLFPGIILSEDFVFCIGYDCSKAIVDRQLREKKG
    SKEKFEIGLYHSARSKALYRNDDELINLIBEVNKNISNNTTKDKDFLFGVYVPD
    DINKIKVTV"
    8867..9184
    /gene="BB0651"
    /complement (9184..9184)
    /note="similar to GB:142023 SP:P4592 PID:1003381
    PID:1222164 PID:1204439 percent identity: 33.33;
    identified by sequence similarity; putative"
    /codon_start=1
    /transl_table=1
    /product="conserved hypothetical protein"
    /protein_id="AAC66994.1"
    /db_xref="GI:2688575"
    /translation="MYSWGQFVFLLOESGNSFLRSLVFPVIAIEFWLIVSPQRK
    BEKNKEMIKNKKKDKVLLIGIGVVKGLDVTIILESPNNEAVPIKNSIDKVL
    EKK"
    9270..11030
    /gene="BB0652"

Query Match      5.9%; Score 110.2; DB 1; Length 13990;
Best Local Similarity 51.4%; Pred. No. 3.1e-14;
Matches 281; Conservative 0; Mismatches 263; Indels 3; Gaps 1;

QY 183 AACTGCGCTGAGATTAAATCGAGCCGATGCTTTGTCGCCGAGTTCGAGCAAG 242
    |||||
DB 10408 ATCTTGCTTAGAGTTAAACGACGCTTTCAGTGTATTAATAATGATGATTGA 10467
QY 243 AGAGCATCTCTCGATCTTGGGAAAAACAATGACACAGGCAATTATCTGACGATGCT 302
    |||||
DB 10468 GAATTAATAGGCGCTACTCTTGAGCTAGACATATTGATCTTGCGCATTTAAAGCTTCGCCG 10527
QY 303 GTGCGTTGGCAATGCTTATGTTTGTATGAGCGATATATTAAGATTGAGAGCGTCATCG 362
    |||||
DB 10528 TTGCTCTTGTAGTTTATTTTGTATATAGTATTATATGATTGAGTGAAGTGTTGAG 10587
QY 363 CT---TGGGAGCGTCTTCTGAAATCTTTGCTTATCTGCGGAGCTCTACGATTTGG 419
    |||||
DB 10588 CTGATTTTCACTGTATTTATTAATGATATTTTAAATTTTAAAGGATAATGTCGCCCTTTA 10647
QY 420 ATGGCGCACTCACCTGTGAGACCTGCTGAGATGTTCTTGTGATAGGGATGAGCGCTAG 479
    |||||
DB 10648 ATTTTACTTTTAACCTTAACAGCATTTGACAGTCTTATTTTGAACAAATGGATGCGCGTTG 10707
QY 480 ATGCAAAATGTTCTTGATTTGAAAGATCCGAGAGCAATTTTATGTCTCAAAAGTCTTA 539
    |||||
DB 10708 ACAATTAATAGTATTTATGAGAGCAATTAAGAGAAATTTAGAGAAAGCAGAAAGATTG 10767
QY 540 AAAAATCTGTAGAAAAGATATACCAAGCTTTTGGAGCATTTTGTGATTTCTTAACCTGA 599
    |||||
DB 10768 AAAATGCTTTTGAAGCTGCTTTTAAAGGCAATTTTATCAATTAATGAGTGAACAAATTA 10827
QY 600 CTACAGTATGGGCTGACGACTTCTTTCTTCTGATACAGGCGCTATTAAGGGGTTTG 659

```

```

Db 10828 CAACATTTATCGGTCGCTTTTAACTCTCTTGGACAGGATTTATCAAGTTTIG 10887
Qy 660 CTTTGACATGATTTTAGGAATTTCTTCTCAATGTTTACGCGCTCTTTTCATGACTAAAT 719
Db 10888 CTTGGCTCTCTCTGTCGTAATTTGGCATCCCTTTTATAGTAGTTTGAATTTTTCAGAT 10947
Qy 720 TTTTCTT 726
Db 10948 TTATTTT 10954

RESULT 24
AE009509/c
LOCUS
DEFINITION
  Brucella melitensis strain 16M chromosome I, section 66 of 195 of
  the complete sequence.
ACCESSION
  AE009509 AE008917
VERSION
  AE009509.1 GI:17982603
KEYWORDS
  Brucella melitensis.
SOURCE
  Brucella melitensis
  Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
  Brucellaceae; Brucella.
REFERENCE
  1 (bases 1 to 11540)
    DelVecchio,V.G., Kapatal,V., Redkar,R.J., Patra,G., Mujer,C.,
    Los,T., Ivanova,N., Anderson,I., Bhattacharyya,A., Lykidis,A.,
    Reznik,G., Jablonski,L., Larsen,N., D'Souza,M., Bernal,A.,
    Mazur,M., Goltzman,E., Selkov,E., Elzer,P.H., Hagius,S.,
    O'Callaghan,D., Letesson,J.-J., Haselkorn,R., Kyripides,N. and
    Overbeek,R.
  The genome sequence of the facultative intracellular pathogen
  Brucella melitensis
  Proc. Natl. Acad. Sci. U.S.A. 99 (1), 443-448 (2002)
  11756688
  2 (bases 1 to 11540)
    DelVecchio,V.G., Redkar,R.J., Patra,G. and Mujer,C.
  Direct Submission
  Submitted (13-NOV-2001) Institute of Molecular Biology and
  Medicine, University of Scranton, Scranton, PA 18510, USA
  3 (bases 1 to 11540)
    Elzer,P.H. and Hagius,S.
  Direct Submission
  Submitted (13-NOV-2001) Department of Veterinary Science, LSU Ag
  Center, 111 Dalrymple Building, Baton Rouge, LA 70803, USA
  4 (bases 1 to 11540)
    Kapatal,V., Los,T., Ivanova,N., Anderson,I., Bhattacharyya,A.,
    Lykidis,A., Reznik,G., Jablonski,L., Larsen,N., D'Souza,M.,
    Bernal,A., Mazur,M., Goltzman,E., Selkov,E., Haselkorn,R.,
    Kyripides,N. and Overbeek,R.
  Direct Submission
  Submitted (13-NOV-2001) Integrated Genomics, Inc., 2201 W. Campbell
  Park Drive, IL 60612, USA
  5 (bases 1 to 11540)
    Letesson,J.-J.
  Direct Submission
  Submitted (13-NOV-2001) Unite de Recherche en Biologie Moleculaire,
  Laboratoire d'immunologie et de Microbiologie, Universite of Namur,
  61 rue de Bruxelles, Namur 5000, Belgium
  6 (bases 1 to 11540)
    O'Callaghan,D.
  Direct Submission
  Submitted (13-NOV-2001) Faculte de Medecine, INSERM U431, Avenue
  Kennedy, Nimes 30900, France
  Location/Qualifiers
  1..11540
    /organism="Brucella melitensis"
    /strain="16M"
    /db_xref="taxon:29459"
    /chromosome="I"
  79..975
    /gene="BMEI0674"
  79..975

FEATURES
  source
    gene
    CDS

  /gene="BMEI0674"
  /codon_start=1
  /transl_table=11
  /product="SULFATE TRANSPORT SYSTEM PERMEASE PROTEIN CYST"
  /protein_id="AAL51855.1"
  /db_xref="GI:17982604"
  /translation="MARKSRRGVMPLSPCFWKSVSLLSMKRNLSVLPGFGLTGLCTLV
  YLTIVLPLMLKTSLSIGWDFWNIVTSERALTATYRTVSAATATLNGFLGLL
  LAWLTRYDPFGKRLIDAAVDLPALTAAGLALVTLFPAONGWFGVLPGLTKVAY
  APLGIMAFPTSPFVTVQPVLEDMADVEAARSLGATSMQTTHITWPTIFPA
  FLASLSPARSLGEGFSIVFISGNLPFQTEVTALLVIRLDEYDYPAAAAALAPVMLV
  MAFFMLATVNLIQARRLYAER"
  962..1792
  /gene="BMEI0675"
  962..1792
  /codon_start=1
  /transl_table=11
  /product="SULFATE TRANSPORT SYSTEM PERMEASE PROTEIN CYSW"
  /protein_id="AAL51856.1"
  /db_xref="GI:17982605"
  /translation="MQNADAALQARNTPVRNLIWLAAGLSVLCIGAPLAVIFSAYF
  SKGIHVPFSSILKPDTLHAYMLVTLTAIVVVPINMAGVCVAMLVTKFRPGRLLIT
  FVPIPSVPIVAGVTYLFYLSQGLLGPFLAHDIKIMETVPAIFLVSLFVTSPPFA
  RELIPLMEVQGSDEEAAALTGASAMQTFYVTLNPKWALLYGAVLCNARVMGEFGA
  TSVVSGAIRGKTNTLPLQVELLFNDVNVAGAFAPAAASTALIALTLVLKVFLEKQHS
  "
  complement(1838..2083)
  /gene="BMEI0676"
  complement(1838..2083)
  /gene="BMEI0676"
  /codon_start=1
  /transl_table=11
  /product="hypothetical protein"
  /protein_id="AAL51857.1"
  /db_xref="GI:17982606"
  /translation="MVASRNDYMSFAKAEALSHVWGSGFVDLGHAGHINVASGFGHW
  PDGAILASSLHREPAPVNPPLGGLPAPRPLPGWAF"
  complement(2129..2443)
  /gene="BMEI0677"
  complement(2129..2443)
  /gene="BMEI0677"
  /codon_start=1
  /transl_table=11
  /product="hypothetical protein"
  /protein_id="AAL51858.1"
  /db_xref="GI:17982607"
  /translation="MPATLLIPGYKSGEAGHQRQWLHDDPSALLVQDDWHYPVLSD
  WMHMLBATLAENPGAVLVAHSLGCVLVAHLSRPAAPAAHVAGALLVAPADAETMARRDS
  RF"
  complement(2670..3641)
  /gene="BMEI0678"
  complement(2670..3641)
  /gene="BMEI0678"
  /codon_start=1
  /transl_table=11
  /product="LOW PH-INDUCED PROTEIN A"
  /protein_id="AAL51859.1"
  /db_xref="GI:17982608"
  /translation="MKADYINWPVIGICAVGSANWLLYRLRSISLDDVLSLYAIRM
  HHWLATASALLAYSLSLADYRLALLHLRFKISLWLFALCSFTYALSHNIGASVSG
  AVSYRAYSQSGMPGSKIAVLIAFCSPFTVLGVIITSSVVLLEPHVLMRPNELTFA
  VSVIVAILMFLFVLLYIFGSMRLPLQIGSFRLVPRLPVLAQQLVAPLELTGAAG
  IIVFALPEAGNPGFLIILGIFLYSFAALISHAPGGIGVLELVPLTGLPMDQADVLA
  ALIVFRLYLLIIPFASLVLVPLVLFKRSQFVLRWNEKQHQEQPGKSR"
  complement(3641..5500)
  /gene="BMEI0679"
  complement(3641..5500)
  /gene="BMEI0679"
  /codon_start=1
  /transl_table=11
  /product="POTASSIUM/PROTON ANTIporter ROSB"
  /protein_id="AAL51860.1"

```

```

/db xref="GI:17982609"
/translation="MPHDPPLATITVIGLCIAPIRGATIRPKISPLVGYLLAGVIAG
PTPGFVADDDLLQLAEIGVILMLFPGVGLHPSKLLISVKAIIVPGALAQIAAAAL
GTGLGLADNDVQAGVFGIALSTASTVLLRMQRRLLIDERGIAIGMLIVEDLA
MMVALILPAISVIGRAGMAASDTIVMLGIGIGITLLIVAKVLIIVMLVGR
KJIPMLNVIAQTSGRRELPRIGVIALGVAFGALFGVSPALGAFPGMGWSESEL
SHRAAEESI.3IRDPFAVLPFISVGMFDMSLLKDLPLGLTALILIDKSVAAFPITV
RARRPVSTALTISALAOIGERSPILAGVSNILPPAGDILLAGALISPIPNV
MTIVARERPMLEKRGATNLTDESANTENHIAEPPLHTLTGTTITVGGQIGGR
LAAIGKEVPLVIDSPKISLKEKIGITVAGNASNTEILEANPGVAAHVIAI
PNAFEGRVIALGRANSIGHIVRACSOAEQALDGLADSVIAGEEBIAGMAVAI
GADPSEBSNVHAPIEVPAGKEMPEKAEDATDITVINGGDMKREGVIG"
/complete(5578..8115)
/gene="BMEI0680"
/complete(5578..8115)
/gene="BMEI0680"
/codon_start=1
/transl_table=11
/product="PROTEIN TRANSLOCASE SUBUNIT SECD / PROTEIN
TRANSLOCASE SUBUNIT SECF"
/protein_id="AAL51861.1"
/db xref="GI:17982610"
/translation="MRTSKVNVVALYSVLVIGIITLIPNFTQKQDALPMLPKQV
TGLDRGGSYLVLEVDAAALKKORLTLLDARSKLRADRIQPSIRVGDVAVVTI
PDADQAKAETALRTLIQVNTSGPGTAINDIDVTENDNNQILTLTEAGFNRLDSA
LQOSETIRORVDQVYAEPSIORVSDRIYVQLGDPQALROLIGSTAKSPHMV
ADANPDPEPPGVITMDSKNPAIKYPIEDQVLSGRLLDAGFDPTNPIVSFR
FSLGARGADITTKVNRFPFVLDGKVLAVIEPIETGSSGQISGFTVDIVL
SALLRAGALPAPLTIEERTVDPDGLGDALKMLMGIIIGFLVAVFILLVPGWVI
ANVALIHLTLPFSALISGATLPLGIDAIKMGIMGIIIGFLVAVFILLVPGWVI
AMALDKGHSARATVDANVTTLTATLITLIFGTGVRGFAVTMGVLSIATILTF
LVNIMAMFVRKKIKVLIHEPFLKVPETHNTRPMNAPRIGIVSIVLSIATILTF
KPLNTGIIIFKGGIOAEISTQPADLQLEAKGALNLGEVALQTSRPNQVIRQR
OEGBEEOJPAINKREAVTELDGVKIEETEVGSELSAQMLAIVLSAVML
FYIWMREPFALGALATLIDTTKIIIGFALQDLPNLIAALMLITIGYVNDVAV
VYDRMENMLYSKTLREIIMDSINOVLRVCYTTLCPMPAIMGSAVHNFA
VPMFCIVATSSSIFAPILILLDGMWRHHAHKTAEGLAOK"
/complete(8224..8640)
/gene="BMEI0681"
/complete(8224..8640)
/gene="BMEI0681"
/codon_start=1
/transl_table=11
/product="Hypothetical Protein"
/protein_id="AAL51862.1"
/db xref="GI:17982611"
/translation="MTGRKKTEDROHGMFLVTCVLSFTAFITVDSPSSQAGDALN
OPRGIARQEGSAAPAROOSTPAPQIPRAIIMEAIAAKIKTVMIGDNAAALLTKTA
ELHVSNGRBAEAAQIPQPLALRGHFSARAPPENNA"
/complete(8962..11457)
/gene="BMEI0682"
/complete(8962..11457)
/gene="BMEI0682"
/codon_start=1
/transl_table=11
/product="mechanosensitive ion channel"
/protein_id="AAL51863.1"
/db xref="GI:17982612"
/translation="MAQNDKQONPASQAPAEKPNQQAASADQRPPEVSPVSI
VOOQRPPEDELCKTKINQOLSKSASDELANLKIQDALSKLLETAVAPRPLT
EINTRLEQLGAPSSDQAPAPAIYGERALIAEKANINATIGETETSLAVNRMSAL
IGDMRDLFTKTLSDQVNLSDTISSEVYSAASDOMISLKIIVSWMRFPVTFKLESL
IAAFALAAALVLOFGARFLGAFYRDPVESVSPYSIRLSAVSWSTVISAAGVFL
ATTYFLINYNVLTIDIASLFOSLFIYGVIFHRLAVACISDMQOMLVQVAPR
GHLLAMVTATLSGLDSFGIVNRILISSPLSTMAKSIATVILGLIALIAVFRP
VERDEGAVAMPAPRIPLIMGLILILALSGVIGIABPISQOIVVGAFLVATVM
GFLTKRAISEBOAFASRRIGKAMERHHPBATIDQGLAGILINIVALLIGPIVL
MQLGFQWAEKSTFYKLTMTGFQIGNFSISLMGLSGVLLPLIGTVLTRNQNLMDNSV

```

Query Match 5.9%; Score 109.6; DB 1; Length 11540;
Best Local Similarity 49.8%; Pred. No. 4,3e-14;
Matches 277; Conservative 0; Mismatches 279; Indels 0; Gaps 0;

```

QY 150 TCTCAGGAAATTTTACCACCGTGAAGTGAAGCAAACTCGCTCAAGATTTMAATCTGAG 209
Db 7151 TCAGCGGAGGCTTACCGGTTCAAGATAGCGTGTCTCTCCGGGCTTCGCGGCGAGGG 7092
QY 210 CGATGCTTTTGTCCCGAGGTTCTCAGTGAAGACATCTCTTGTATCTTGGGAAA 269
Db 7091 CCTGCGCCCTCGCTCACCGTCATCGAAGAACGACCGCTGCGCCGAGATCTTGGGGCG 7032
QY 270 AACATGTACAAAGGCATTATCTAGCATGCTGAGTGGCTTGGCAATGCTATTGTTTGA 329
Db 7031 ACCGCATCAAGATGGGCTGATGACAGATATATCGGCTTCTTCGTCGCGGCTTCA 6972
QY 330 TGAAGCTATATTATAGATTGGAGGCGTCATCGCTTCGGAGCTGTTCTTGAATCTTT 389
Db 6971 TCCGTGCTTCTATGATTCGCGGTTGATTCGGAATTCGCGGCTTCGCTCATACCG 6912
QY 390 TCGTATCTGGGCGAGCTTACAGATATTGGATGCGCCACTCACCTTGTACAGACTGCTG 449
Db 6911 TCGTACCTTCTCGGCGCTAAGCGCTGATCGGCGCAACGCTGACGCTGCCGATGACG 6852
QY 450 GGAATGCTTCTGATGGGATGGCCGTAGATGCAAAATGTTCTTATCGAAAGATCC 509
Db 6851 GCATTATCTCGGATCGGATCGCGTGCATGCGAATTTCTCATCAACGAGCGCATCC 6792
QY 510 GAGAGAAATTTTATTGTCTCAAGTCTTAAATAATCTGTAGAAAAAGATATACCAAG 569
Db 6791 GTGAAGAACCCGCAAGGAGACTTGGCGCATGGCGGCGCTCGACAAAGGCTTCACAGCG 6732
QY 570 CTTTGGAGCATTTTGTATTCTAATTGACTGACTAGAGTATTGGCTTCTTCTTCT 629
Db 6731 CTTTGGACCATTCGAGCGGACGCAATGTGACGACGCTGACGCGCAACATCTTCTGTTCC 6672
QY 630 TCTTACATACAGGCGCTATTAAGGTTGCTTGAATTGATTTAGAAATTTCTCT 669
Db 6671 TGTTCGACACGGCCGCGTTCGCGGCTTCGCGCTCACCATGATGCTCGGATGCCATTT 6612
QY 690 CAATGTTACGCTCT 705
Db 6611 CGATGTTACCGATGT 6596

```

RESULT 25
MPUM03/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

MPUM03 315079 bp DNA linear BCT 14-MAY-2001
Mycoplasma pulmonis (strain UAB CTIP) complete genome; segment 3/3.
AL445565 AL445566
AL445565.1 GI:14089942

Mycoplasma pulmonis.
Mycoplasma pulmonis
Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.

REFERENCE
AUTHORS
TITLES

1 (bases 1 to 315079)
Chambaud, I., Heilig, R., Ferris, S., Barbe, V., Samson, D.,
Gallison, F., Moszer, I., Dydyv, K., Wroblewski, H., Viat, A.,
Rocha, E.P.C. and Blanchard, A.
The complete genome sequence of the murine respiratory pathogen
Mycoplasma pulmonis
Nucleic Acids Res. 29 (10), 2145-2153 (2001)

JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLES
JOURNAL

2 (bases 1 to 315079)
Blanchard, A.
Direct Submision
Submitted (16-OCT-2000) Blanchard A., INRA, Centre de Recherche de
Bordeaux, Institut de Biologie Vegetale Moleculaire, 71 avenue
Edouard Belin, BP 81, 33883, Villenave D'Ornon, Cedex, France

FEATURES
source
1..315079
/organism="Mycoplasma pulmonis"
/strain="UAB CTIP"
/db_xref="taxon:2107"
288..2807
/gene="MPU_5280"


```

/translate=4
/product="unknown; predicted coding region"
/protein_id="CAC13707.1"
/db_xref="GI:14089949"
/translation="MEKINNYEDYLKKKDLIKOKISLISLSTLWNSIMILNIV
ILAAAVINVEIKRVGNLPLTRKSPFDNGLTIYASVLVFLFNLALVYPTLMK
WDOYKXANMELEYIKHNDODKYEKEGLDLENLKKTLWLFKNKISFETLKKFLLD
KNA"
10725..11624
/gene="MYPU_5350"
/locus_tag="MYPU_5350"
/feature="MYPU_5350"
/feature_id="1"
/feature_start="1"
/feature_end="11624"
/feature_type="CDS"
/feature_note="no similarity found"
/codon_start=1
/evidence="not experimental"
/translation="MLKNKISIVLEPAOKAKKTLKKKVRISIFMLITFLVFL
FSAIWGLAKIASRPFQISFIVNPFPSKIKENGKEIYVNDWPIVFLITGIFSIING
LMLFAIKWVENKERNIOISLOPKLFEASEBKIDOLKREKEFLFQINSILTS
SRINKKELEIOEKIKCCNTSIDBELYKLOIKHLRSKIFLFPNISILISLWNG
FLASYSIAKSNLESVKIFIVISFIVVSVFLTNLVTLSLAKQIDOLEKIDLM
Query Match 5.8%; Score 108.8; DB 1; Length 315079;
Best Local Similarity 49.5%; Pred. No. 4.6e-16;
Matches 281; Conservative 0; Mismatches 287; Indels 0; Gaps 0;
Query 153 CAGGAATTTATCCACCGTGAAGTGAACAACCTGCCTCAGATTTAAATCTGAGAGA 212
Db 143811 CTGTAATTTACTGCTCTTACGTAACAACCTGCCGTGATGAACCTTGGTTCAG 143852
Query 213 TGTCTTTTGTCCGAGGTTCTCAGTGAAGAAGCATCTTCTGATTTGGGAAAAAC 272
Db 143851 GTGATTAACCTTAAGAGCTTTATCTTCAGATTTGTGAAGCTACTTTAGAACTCAT 143792
Query 273 AATGTACACAAGCATTTCTCAGATCTGCTGGCAATGCTTATGTTTGAATGA 332
Db 143791 CATTTAACTATGCAATTAATTCAGGAATTAATTTCAATTTATGTAATTCCTTAATTA 143732
Query 333 GCGATATTATAGATTTGAGAGCGCATCGCTTCGGAGAGCTTCTTCTGATCTTTTGC 392
Db 143731 TTTACACTATGCTGTTTCTGATTAATGAGAACTCAACAGCTCAATAGCGGCTT 143672
Query 393 TTATCTGGGAGCTCTACAGATTTGATGCGGCACTCACTTTCAGAGACTCGTGGGA 452
Db 143671 TTTCACTAACTGTTTCTGATTAATGAGAACTCAACAGCTCAATAGCGGCTT 143612
Query 453 TTGTTCTTGTGATGAGGATGCGCGCATGCAAAAGTCTTGTATGGAAGAAATCCGAG 512
Db 143611 TAGTAAATAGCTAATAGATTAATGATGCGCAATTAATTAATTAATTAATTAATTA 143552
Query 513 AGGAATTTTATGCTCTAAGTCTTAAATAATCTGTAGAAATAAGATATACCAAGGCTT 572
Db 143551 AGCACTCAATATGAGAAATGCTTAAATAAGCTATTAATAATGCTATGATCTCTGC 143492
Query 573 TTGAGCAATTTTGAATCTTAACTGACTACAGATTTGCGCTCAGACACTCTTTTCTTC 632
Db 143491 TTTCAACAATTTTGAGCAATTAATTAATGATGAGAAATTTCTTTTCTTCT 143432
Query 633 TAGATACAGGCGCTTAAAGGCTTTGCTTGAACATGATTTAGGAATTTTCTCTCA 692
Db 143431 TTGATACAAAGCTCATTAAGGCTTGAACATGATTTCTGAGAAATTAATTAATTT 143372
Query 693 TGTATACGCGCTTTTCACTAGCTAAAT 720
Db 143371 TGTATGATGTTAATTAATCACTAGAT 143344

```

RESULT 26
AX084084 1848 bp DNA linear PAT 28-FEB-2001
LOCUS

```

DEFINITION Sequence 119 from Patent WO0111033.
ACCESSION AX084084
VERSION AX084084.1 GI:13185634
KEYWORDS
SOURCE
ORGANISM
REFERENCE
1 (bases 1 to 1848)
AUTHORS Choyan, L.E., Heseler, P.B. and Reich, K.A.
TITLE Essential bacteria genes and genome scanning in Haemophilus
JOURNAL Influenzae
ABSTRACT Patent: WO 0111033-A 119 15-FEB-2001;
ABBOTT LABORATORIES (US)
FEATURES
source
1..1848
location/Qualifiers
CDS
1..1848
/organism="Haemophilus influenzae"
/db_xref="taxon:727"
/translation="MLNRPYKMLNIVFIYVAGILYSLPNTYGEDPAVQISCTGQE
ANTSVIGVQVQVLTAKTNMLPTKSVILENGSILAFINTDQILAKOKLARLNNTTA
LNLAPATPMLFPGANPMKGLDVGVRFLMDMNAITLYKREODLSLGEJLARK
EKIOYTAIENHFGTLITLAVSOKAKERIIRHPLDITDEPDADISLGLSTAA
LNEADIALEONTLIRKRVATGVAENIORGAKERYIEIYVGGDETHIINSSGLD
ATLERIYNQNTAALISRNMLPADSEVYDQGHVLPKRAVGGDETHIINSSGLD
QHSSTPQVSVLDSGGEIWSQTKTKYKPMATLYVERKNDKDKENKTLLEKHE
VINVAITQGRFGSNFOITGVDSIAEAMHSTLKSGLALPIQIVERTIGSLGQON
VEGINASLWGLVAVIAFMLFYKMGVSLPALVINIVLWGLSILFGLATLSMGI
AGIVTLGMSVAVNLIPEIKERIRNGSIOQALINEGNGAFSTIFDNLTLTLTAI
ILVAVGTGPIQGFATILSLGVALSMETATITGRALVNALYKQJLKLII"
BASE COUNT 560 a 321 c 388 g 579 t
ORIGIN
Query Match 5.8%; Score 108.2; DB 6; Length 1848;
Best Local Similarity 51.4%; Pred. No. 1.1e-13;
Matches 276; Conservative 0; Mismatches 258; Indels 3; Gaps 1;
Query 173 GAACTGAGCAAACTGCTCAGATTTAAATCTGAGAGATGCTTTTCCGAGGT 232
Db 1249 GAAGCATATATCTTTCTACTATTAATAATCTGAGCATTAATGACCAATTCGAAT 1308
Query 233 CTGATGAAGAAGCATCTCTTCTGATCTTGGGAAAAACAAGTACACAAGCATATC 292
Db 1309 GTTGAAAGCAAGCAATGCGCCATCTTAAGTGGCGAAACGTAAGCAAGGATTAAT 1368
Query 293 TCAGATGCTGTGGCTTGGCAATGCTTAATGTTTGAAGCGTATTAATTAATTTGA 352
Db 1369 GCGAGTCTTTGGGATTAATGCTGTTATGCTTTATGTTTATCTAATAATGTTT 1428
Query 353 GGCCTCATCG--TTGGGAGCTGTTCTTGAATCTTTTGTCTTATCTGGGAGCTCTA 409
Db 1429 GGTGTGATGGAAGTTTGAACCTTATTAATATCGTATTAATCTTGGGATTAATGCT 1488
Query 410 CAGATTTGATGAGGCACTCACTTGTCAAGACCTGCGTGGAGTGTCTTGAATGGGG 469
Db 1489 ATTTTACCGGCGGACACTTCAATGCCGGGATATGCGGGATCGTTTAACTTTAGGT 1548
Query 470 ATGGCCGATGAGAAATGTTCTTGTATTCGAAAGAAATCCGAGAGAAATTTTATGCT 529
Db 1549 ATGTGATGATGAGAAATGATTTTGAAGATTAATAAGAAATTCGTAATGCT 1608
Query 530 CAAAGCTTAAATAATCTGTAGAAAAGATATACCAAGCTTTTGAAGCATTTTGTAT 589
Db 1609 CGTTCAATTTGCAAGCATTAATGAAGGTTATACCGCGCATTTACTTATTTTGTAT 1668
Query 590 TCTAATCTGATACAGATTTGCGCTCAGACACTTCTTTTCTTCTAGATACAGGCGCTAT 649
Db 1669 GCAAACTTAACCAATCTTAACCGCAATTAATTTCTATACGCGGTAGGACAGGCCCAAT 1728

```


Qy 650 AAAGGTTTGGCTTGCACATGATTTTAGAATTTCTCTTCAATGTTTACGGCTCTT 706
 Db 1729 CAAGGTTTGGCAATTCAGCTTTCACCTTGTGTGGCATTTCTTATCTGTATCCCGGATT 1785

RESULT 27
 AX395145 1848 bp DNA linear PAT 18-MAY-2002
 LOCUS
 DEFINITION Sequence 119 from Patent WO0218601.
 AX395145
 ACCESSION
 VERSION
 AX395145.1 GI:21066198
 KEYWORDS
 SOURCE
 ORGANISM
 Haemophilus influenzae.
 Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
 Haemophilus.

REFERENCE
 1 Chovan, L.E., Hessler, P.E. and Reich, K.A.
 AUTHORS
 TITLE Essential bacteria genes and genome scanning in I Haemophilus inf
 JOURNAL Patent: WO 0218601-A 119 07-MAR-2002;
 Abbott Laboratories (US)
 FEATURES
 source
 1. 1848
 /organism="Haemophilus influenzae"
 /db_xref="taxon:727"
 1. >1848
 /note="HI-0240"
 /codon_start=1
 /transl_table=1
 /protein_id="CAD32145.1"
 /db_xref="GI:21066198"
 /translation="MLNRYPLWKNLWIFVAIGILYSLNPYIGEDPAVOISGTRGOE
 ANTSLVGQVODVLKTNLPKTSIVLENGSILARTNTDDOLLAKDKIAERLGNNTYA
 LNLAPATPAWLSMFGANPMKGLDRCGVRELFMEVDNATLVKQEOLOQSLRGELEK
 EKIOYTKNTEHSGTLTLTANVSORAKAERIIRQLHPTLDITEPDADSINGLSTAA
 LNEARDLAEONULIRKRVAGLVAERIVQROGAERIVILPQVQDTARAKELIGAT
 ATLEFRIVQNTVDALISRNMLPADSEVKYDROGHPVALFKRVLGGEHTINSSGLD
 QHSPTQVTLDSGEIEINSQTKYKKPMATLYVYKDKNGKKBENGKTIIEKHEE
 VINVATTCGRFSNQITGDSIAEHNLSFTLKSGLAIPIQIVERTIGPSLGAQN
 VEQGINASLWGLVAVIAPMLFYKMGVFIASFAVINIVLLGLMSILPGATLSMPCI
 AGIVLTGMSVDANVLIFERIKERNGRSIQQAINEGYNGAFISIDANITLTITAI
 ILYAVGTGPIGGFAITLSLGVASMFATITGTRALNVALYGGKOLKLLI"

BASE COUNT 560 a 321 c 388 g 579 t

Query Match
 Best Local Similarity 5.8%; Score 108.2; DB 6; Length 1848;
 Matches 276; Conservative 0; Mismatches 250; Indels 3; Gaps 1;

Qy 173 GAAGTGAGCAAACTCGCTCAGATTAAATCTGGAGCGATGCTTTTGTCCCGAGGTT 232
 Db 1249 GAAGCACATAATCTTTCTACCTTATTGAATCTGTGCAATTAATGCACCAATTCAAATT 1308

Qy 233 CTCAGTGAAGAGACATCTTCTGATCTTGGGAAAAACAATGTACACAGGCATATC 292
 Db 1309 GTTGAAGAACGCACAAATTGGCCCATCATTAGTGTGGCGCAAACTAGAGCAGCGGATTAAT 1368

Qy 293 TCAGCATGCTGTGGCTGCGCAATGCTTATGTTTGTGAGCGGTATATATAGATTGGA 352
 Db 1369 CGGAGCTTTGGGGATTAGTGTCTGTTATGTCCTTTATGCTTTTACTACAAAATGTTT 1428

Qy 353 GCGCGTCATCGC---TTCGGGAGCTGTTCTTCTGAATCTTTTCTTATCTGGGCGACTCTA 409
 Db 1429 GGTGTGATTGCAAGTTTTCGACTTGTATTAATATCGTATTACTTGTGGGATTATGTCT 1488

Qy 410 CAGTATTGGATGCGCCACTACCTTGTGACGACTCGCTGGGANTGTTCTTGTCTATGGG 469
 Db 1489 ATTTTACCGCGCGCACACTTTTCAATCGCGGTATTGCGGGTATCGTTTAACTTTAGGT 1548

Qy 470 ATGCGCGTAGATGCAATGCTTCTGATTCGAAAGAACCCGAGAGAAATTTTATGTCT 529
 Db 1549 ATGTCAGTAGATGCGAATGTATGATTTTGTGAACGATATAAAGAGAAATTCGTAATGCT 1608

Qy 530 CAAAGTCTTAAAAAATCTGTAGAAAAAGGATATACCAAGGCTTTTGGAGCCATTTTGTAT 589
 Db 1609 CGTTCAATTGAGCAAGCCATTAAATGAAGGTTATAAGCGCGCATTTACTTCTATTATTGAT 1668

Qy 590 TCTAACTTGACTACAGTATTGGCCTCAGCACTTCTTTCTTCCTAGATACAGGCGCTATT 649
 Db 1669 GCAAACTTAAACCAATCTTAAACCGCAATTTCTATACGCGGTAGGACAGGCCCAATT 1728

Qy 650 AAAGGTTTGGCTTGGACATGATTTTAGGAATTTCTCTCAATGTTTACGGCTCTT 706
 Db 1729 CAAGGTTTGGCATACGCTTTCACCTTGTGGTGTCTATGTTTACCGCGATT 1785

RESULT 28
 U32710/c 13818 bp DNA linear BCT 29-MAY-1998
 LOCUS
 DEFINITION Haemophilus influenzae Rd section 25 of 163 of the complete genome.
 U32710 L42023
 ACCESSION
 U32710.1 GI:1573200
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Haemophilus influenzae Rd.
 Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
 Haemophilus.

REFERENCE
 1 (bases 1 to 13818)
 AUTHORS
 Fleischmann, R.D., Adams, M.D., White, O., Clayton, R.A.,
 Kirkness, E.F., Kerlavage, A.R., Bult, C.J., Tomb, J., Dougherty, B.A.,
 Merrick, J.M., McKenney, K., Sutton, G.G., FitzHugh, W., Fields, C.A.,
 Gocayne, J.D., Scott, J.D., Shirley, R., Liu, L.I., Glodek, A.,
 Kelley, J.M., Weidman, J.F., Phillips, C.A., Spriggs, T., Hedblom, E.,
 Cotton, M.D., Utterback, T., Hanna, M.C., Nguyen, D.T., Saudek, D.M.,
 Brandon, R.C., Fine, L.D., Fritchman, J.L., Fuhrmann, J.L.,
 Geoghegan, N.S., Gnehm, C.L., McDonald, L.A., Small, K.V., Fraser, C.M.,
 Smith, H.O. and Venter, J.C.
 TITLE Whole-genome random sequencing and assembly of Haemophilus
 influenzae Rd
 JOURNAL Science 269 (5223), 496-512 (1995)
 MEDLINE 95350630
 PUBMED 7542800

REFERENCE
 2 (bases 1 to 13818)
 AUTHORS
 Tatusov, R.L., Muehlegan, A.R., Bork, P., Brown, N.P., Hayes, W.S.,
 Borodovsky, M., Rudd, K.E. and Koonin, E.V.
 TITLE Metabolism and evolution of Haemophilus influenzae deduced from a
 whole-genome comparison with Escherichia coli
 JOURNAL Curr. Biol. 6 (3), 279-291 (1996)
 MEDLINE 96398784
 PUBMED 8805245

REFERENCE
 3 (bases 1 to 13818)
 AUTHORS
 White, O., Clayton, R.A., Kerlavage, A.R. and Fleischmann, R.D.
 TITLE Direct Submission
 JOURNAL Submitted (25-JUL-1995) The Institute for Genomic Research, 9712
 Medical Center Dr, Rockville, MD 20850, USA

REFERENCE
 4 (bases 1 to 13818)
 AUTHORS
 White, O., Clayton, R.A., Kerlavage, A.R. and Fleischmann, R.D.
 TITLE Direct Submission
 JOURNAL Submitted (27-SEP-1997) The Institute for Genomic Research, 9712
 Medical Center Dr, Rockville, MD 20850, USA

REMARK
 The H. influenzae sequence has been updated by R. Fleischmann. New
 database matches have been assigned, product names have been
 improved, and a number of frame shifts have been corrected. We
 gratefully acknowledge the work of Tatusov et. al. We have
 incorporated their annotation into the /notes fields of the
 corresponding H. influenzae genes

REFERENCE
 5 (bases 1 to 13818)
 AUTHORS
 White, O., Clayton, R.A., Kerlavage, A.R., Fleischmann, R.D.,
 Peterson, J., Hickey, E., Dodson, R. and Gwinn, M.
 TITLE Direct Submission
 JOURNAL Submitted (28-MAY-1998) The Institute for Genomic Research, 9712
 Medical Center Dr, Rockville, MD 20850, USA

REMARK
 The whole genome was shifted by 588 nucleotides for a new start
 On Sep 30, 1996 this sequence version replaced gi:1220911.
 FEATURES
 Location/Qualifiers

```

source      1. .13818
            /organism="Haemophilus influenzae Rd"
            /db_xref="taxon:71421"
            /complement(134. .343)
gene        /gene="HI0235"
            /complement(134. .343)
CDS         /gene="HI0235"
            /note="similar to GB:I29458 SP:P36675 PID:473421 percent
            identity: 73.81; identified by sequence similarity;
            putative"
            /codon_start=1
            /transl_table=11
            /product="conserved hypothetical protein"
            /protein_id="AAC21905.1"
            /db_xref="GI:1573201"
            /translation="MRKOKSAVENETVYLTFRGVIKDNVAALLQDKLFRQREKKR
            KGGSYORAKHPGKIFKPDYKF"
            /complement(411. .761)
gene        /gene="HI0236"
            /complement(411. .761)
CDS         /gene="HI0236"
            /note="similar to PID:1805555 percent identity: 45.74;
            identified by sequence similarity; putative"
            /codon_start=1
            /transl_table=11
            /product="arsenate reductase, putative"
            /protein_id="AAC21906.1"
            /db_xref="GI:1573202"
            /translation="MSVILYHNPHCKSRRETLLENKGIQPIIEYLQKOYSVNEIQ
            STAKKIGIDVRQMWRTKDELYKSLNDNLDSQAELEKAISSHALIERPIVINDK
            AKGRPEPTVLEIL"
            /complement(840. .1868)
gene        /gene="HI0237"
            /note="This region contains an authentic frame shift and
            is not the result of a sequencing artifact; similar to
            GB:U00096 SP:P77406 PID:1621505 PID:1788838 percent
            identity: 55.75; identified by sequence similarity;
            putative"
            /pseudo
            /complement(2003. .2980)
gene        /gene="HI0239"
            /complement(2003. .2980)
CDS         /gene="HI0239"
            /note="similar to SP:P19674 GB:X56175 PID:581231 GB:U00096
            PID:1773093 percent identity: 48.00; identified by
            sequence similarity; putative"
            /codon_start=1
            /transl_table=11
            /product="protein-export membrane protein (secf)"
            /protein_id="AAC21907.1"
            /db_xref="GI:1573204"
            /translation="MMKLFYKDXDGHFIREINGIKLPPLTERMKVKKLYISALLIM
            VISLFFITKGFNMGDLFTGVVDEFTQSANLEQSRSLKHNGESSPIVQTTGSVQ
            DWSIRLPASNNDSITGHEVSKMLQNVKDIQISIEFVGVSELAQAGVATLATL
            AMVLIVGSRFPEMRKRGCSIASLAHVITILGVPSAIOEITDITFVAALISVGVGIN
            DSIVFEDRVENFRKIRLDITIDISLQVTSRTITTSVTLIVVMALFFFGG8BI
            HNFSLALVIGIGETYSIFVAIAVAVDGLRREHMLPPVDXEIDELP"
            /complement(2988. .4838)
gene        /gene="HI0240"
            /complement(2988. .4838)
CDS         /gene="HI0240"
            /note="similar to SP:P19673 GB:S68715 GB:X56175 PID:581230
            PID:545175 percent identity: 59.74; identified by sequence
            similarity; putative"
            /codon_start=1
            /transl_table=11
            /product="protein-export membrane protein (secf)"
            /protein_id="AAC21908.1"
            /db_xref="GI:1573205"
            /translation="MLNRYPLMKLMLAVIFVAILGILSLPNITXGEDPAVOISGTRGQ
            ANTSVGGVQDVLTNNLPKTSIVLENGSILARTNTDDQLADKLAELKGNVYTA
            LNLAPFAWLSMGANPMKMGKIDLRGVFIMVDNNATILVROEQLDLSLGELEK
            EKIQYIAKTEHFGTILITLANSQRAKARIIRQLHPTIDITRPDSDSINIGLSTA

```

```

INAEADLAIEONLTLRKVAELGVAEAVIQRQGAERIVIEIPGVQDTPARAKEILGAT
ATLERIVNONTVDATSRNMLPADSEVYKROGHPVLPRAVIGGEHIIINSGLD
OHSSTPYSTLIDSGEINSGOTTKYKKMALTIVYKNGKXKDNKXTLEKHEE
VINVATTIGRGRGNSNQTIGVDSIAEHNLSLILSGALIAIPQIYEEITGIBSICAN
VEQGINASLWGLVAVIAFMFLFYKMFQVIAFPALVINLVGLMSLILPGATLSMPGI
AGIVLTIGMSVDANVILPERIKERIINRSIQOALINBYNAGFISIPANITLITAI
ILVAGTPIQGFALTSLGVAISMFATITGRALVNLVYGGKQKLLI"
            /complement(4908. .5201)
gene        /gene="HI0241"
            /complement(4908. .5201)
CDS         /gene="HI0241"
            /note="similar to PID:599589 percent identity: 62.37;
            identified by sequence similarity; putative"
            /codon_start=1
            /transl_table=11
            /product="conserved hypothetical protein"
            /protein_id="AAC21909.1"
            /db_xref="GI:1573206"
            /translation="MEQSPMSSTLFPIVIFGLIFPMATYRPOAKENKHKMLSELAK
            GREVLTAGVIGIKITKTEGSDSIVIALNDTTEITINNYIVSVLPKSLKSL"
            /complement(5309. .5530)
gene        /gene="HI0242"
            /complement(5309. .5530)
CDS         /gene="HI0242"
            /note="similar to GB:L77117 SP:Q58397 PID:1591652 percent
            identity: 35.71; identified by sequence similarity;
            putative"
            /codon_start=1
            /transl_table=11
            /product="conserved hypothetical protein"
            /protein_id="AAC21910.1"
            /db_xref="GI:1573207"
            /translation="MKYOLNLTALRCPIPLSAKKALKNDLMLILNLSAVEN
            RSIAPBNSVALVEQYVSRKPIYILKK"
            /complement(5527. .6045)
gene        /gene="HI0243"
            /complement(5527. .6045)
CDS         /gene="HI0243"
            /note="hypothetical protein; identified by Genemark;
            putative"
            /codon_start=1
            /transl_table=11
            /product="H. influenzae predicted coding region HI0243"
            /protein_id="AAC21913.1"
            /db_xref="GI:1573208"
            /translation="MQILRPQGFATMNEPIEMLYACHSVKAPRCQSLTPLYEKHG
            YTOAVLNDVEQILSTFNRAAPLHDDDEIDFPQVAVAPQVTSIDLEKQHEYLHE
            MNMALSVQLEELISEQRDIDKHLIERFIQGYDRHIALEPLFEMGRECTSLDILTEM
            GKMSARQVKE"
            /complement(6110. .7258)
gene        /gene="HI0244"
            /complement(6110. .7258)
CDS         /gene="HI0244"
            /note="similar to GB:M63939 SP:P19675 GB:S68715 GB:X56175
            PID:147966 percent identity: 81.60; identified by sequence
            similarity; putative"
            /codon_start=1
            /transl_table=11
            /product="trna-guanine transglycosylase (tgc)"
            /protein_id="AAC21911.1"
            /db_xref="GI:1573209"
            /translation="MKYELDKTSGNARRGLVFERPQGFSTVETPAFMFVGTYGVKG
            MPEBPARAGAILIGNTPELMLRPGCEVMRNGDHDHMOHMRPILTDSGVOVSL
            GKURKTEBGFVQNFINGERIFLSPKSMETQYDLSITWIFDECTYPRATFPAK
            KSMESLIRMAKRSRDFDELGNKNALFGIIOGVFEELKVSLEGLVNIIGFQVYAGG
            LAVGEPEKDMRILLEYICPOIPADKXRYLIMGVPEDELVEGVARGIDMCPWPTNNA
            RNHGLFVTGIVAIRNAKYRDTSPDPEDCDCTYCNVTKAYLYHIDKCGEILIGARLN
            TTHNLRYVORLMAEITRQATEDDRFDPFVVEFYARMGKVPVPQLADKS"
            /complement(7602. .8693)
gene        /gene="HI0245"

```

Query Match

5.4%; Score 108.2; DB 1; length 13818;

Best local Similarity 51.4%; Pred. No. 8.7e-14;


```
/product="glutathione dependent formaldehyde  
dehydrogenase"  
/protein_id="BAB48368.1"  
/db_xref="GI:14021757"  
/translation="MKTRAAVAAVAAAGKLEIMEVDLDGPEGEVLEIKATGICHTDE  
FTRASADPEELFPAILGHGAGVAVDGKVTGSKGDHVIPLYTEPCRCQPCSLRK  
TNICTAIRATOGGIMPDSSRPSIGDKIENHMGCTGSENFLPEIATAKNPAP  
PDKICTYGC3VTTGICAVINTAYEQGATTVFGIGIGLGNFVLQKLAGADMTICVD  
LNNDKKAAGERFQHTFVNPKEDIDGVPHLVNMTKRGADQIGADYTPCTGNTKM  
ROALEASHR3MGKSVIVGAGAGQEISTRPFQVLTGTMTWGTAFGAGRGTDVPRIVD  
WYMDGKIQIDPMITHLKLEIDINKGFDLMHEKGSIRSVVY"  
8113. .8787  
/gene="m1r0874"  
8113. .8787  
/gene="m1r0874"  
/note="hypothetical protein"  
/codon_start=1  
/transl_table=1  
/protein_id="BAB48369.1"  
/db_xref="GI:14021758"  
/translation="MRASRSVASHFTEASHTPTDNPATISVONWEKMAEKLHPR  
DNGI.PKESASFAGITLVCACTSKRVKVGQJLAHNNACCTCKMKEGAI.FSVVAA  
GTGDDVTVENGDKLVNDASALIQHACTCGVHMGVPERDHA.FKGLTIHERFVE  
DGWSPGFIAFVSSIRESVDPKRMDGIRAOQLRTIGLEPCVLCNPLGMDYIATWTAK  
SGALPA"  
8818. .9318  
/gene="m1r0875"  
8818. .9318  
/gene="m1r0875"  
/note="hypothetical protein"  
/codon_start=1  
/transl_table=1  
/protein_id="BAB48370.1"  
/db_xref="GI:14021759"  
/translation="MPAPVRFGEIMPAITVVDADAPYKAEKVEKLEAVVTVV  
SNGLRSPEDPMIRNVVSKGADADADIVENKSNVVDATDIPLAIRAVLGAHPL  
GPTGRPPTPETIGAVAMRDLKQHLRTGSKGNASFAAQDSRFLGELDRILRAL  
KSATPG"  
9338. .10027  
/gene="m1r0876"  
9338. .10027  
/gene="m1r0876"  
/note="hypothetical protein"  
/codon_start=1  
/transl_table=1  
/protein_id="BAB48371.1"  
/db_xref="GI:14021760"  
/translation="METPAAEFPITPRVORHMLFAMACFGALAFVALLHAPLAY  
SIGANAFPAVYVILVVSOMPKFTGKYSKNAARDQVLIPIAVTLVVVAVASLEL  
LINOKRGNSEVLEGFALSTPLGFTTHAAMALHAYVMMDDADAVDAEIKKICPUG  
LTFPGDR.PEGMDLIFYSAVIGMTAQADTNISITTHRCVVLVHSILSFFNFIVAA  
AVMLAVSL3P"  
10125. .11006  
/gene="m1r0877"  
10125. .11006  
/gene="m1r0877"  
/note="hypothetical protein"  
/codon_start=1  
/transl_table=1  
/protein_id="BAB48372.1"  
/db_xref="GI:14021761"  
/translation="MQSVARSQSGATGDADVSSDSTLIYQSRRTLTHIMASLFF  
MLSLGQIFNARPOLYIGKQSGEGYNNITFIEIGVNDVADPRGTETLFGHFPDTGVL  
GWSGPPNETNARPGPSWATIPSYVDLGTAVVHFVFFMTITLLIWLVASLVNGHRI  
RDLAPRLDDELRLPRDIDVHARLKFHTRENTLOKRAYGVGVFLPLMIITGLAMS  
PSMNSVLPFLNEVYGGROTARTIHTFTMLLVGFIIHMLLAAGINELRSIITGM  
YRTDPPHDDKXTERSA"  
10975. .11817  
/gene="m1r0879"  
10975. .11817  
/gene="m1r0879"  
/note="hypothetical protein"  
/codon_start=1
```

```
/transl_table=1  
/protein_id="BAB48373.1"  
/db_xref="GI:14021762"  
/translation="MTTRPRGVPMMAKFEISPNFOISRKKFLTAASLGASGIMLSGC  
DDFDSOLSGSGLRSFLEMANOLTVRAOQLLAGRALAPNEFEADIRPORNGVYAP  
LADDTYKLSANSLRLENSGLVLEKPLSLSRQOLNMFSEIRQIRHDEWMSCLAK  
WGTPTSLVDAAVYAPQARYVFPCLDITDLSGDIKYVCTIIDLRHQITLAIY  
GNGKPLPVENAPLARYVERQGLKMPRYLKIELVDFANVGGRGVWMDNGIDW  
YGGI"  
complement(11866. .12867)  
/gene="m110880"  
complement(11866. .12867)  
/gene="m110880"  
/note="unknown protein"  
/codon_start=1  
/transl_table=1  
/protein_id="BAB48374.1"  
/db_xref="GI:14021763"  
/translation="MGELIMSAPVAVGLTSKNRIAAEDVLMRRDYFADGVTRGEAE  
L.FALDQADKCAEWAPFEVAVTVDIYVHQEKSQVTSODNADMIVRTISRDGMDSR  
TELELIVHLEKARSPSKLCYALAEQVHAHYIDKGP.LMLGALVPGI.IAKAEVELL  
RLIVAHGSDNIAITRGEAEVLFRIEOTAQAKDPSMNDL.FVKAIAFVWCSAGYE
```

Query Match 5.7%; Score 106.4; DB 1; Length 349619;
Best Local Similarity 48.2%; Pred. No. 1.5e-13;
Matches 299; Conservative 0; Mismatches 321; Indels 0; Gaps 0;

QY	150	TCTCAGGGAATTTTACCACCGTGAAAGTGACAAACTCGCTCAGATTAAATCTGGAG	209
Db	202322	TCTCAGGCAATTTTACCACCGCCAGAGCGCCAAATGATCTGGCCGCTTTGGCCGCGTG	202263
QY	210	CGATGCTTTTGTTCGCCAGGTCCTCAGTGAAGACGATCTCTTGATCTGGGAAA	269
Db	202262	CGTGCCGCGAGCGTGACCGGTATGAGAGACGACGTGGTCCGGCTTGGCCAGG	202203
QY	270	AACAATGTACCAAGCATATCTCAGACATGCTGTGGCTTGGCAATCTATTGTTTGA	329
Db	202202	ATTGATTCATGCCCGGAAGTGGCGGCATCATCGGTTGATCTCGTGTGGCTTCA	202143
QY	330	TGAGCGTATATTAATTAATTTGAGGCGCATCGCTTCGGAGCTGTTCTTGAATCTTT	389
Db	202142	TGTTCCGCGCTACCGGCTTCCTCGGTTCTCGCCAAACATCGCGCTGGCCGTCACGTGG	202083
QY	390	TGCTTATCTGGGACGCTACAGATTTTGGATGGCGACACCTGTGAGACTGCGT	449
Db	202082	CGATGATCGTCGCGCTGTTTCGCTGCTTGGGCGAGCGCTGACCTGCGGATTTGCGG	202023
QY	450	GGATTTGTTCTGTATGGGATGCGGTAGATCAAAATGTTCTTGTATTCGAAAGATCC	509
Db	202022	GTAICGTCGACCATCGGATGCGGATGCGGTGATTCGAAAGTGTGATCAGAAAGCATCC	201963
QY	510	GAGAGAAATTTTATGTCTCAAAATGTTTAAATAATCTGAGAAAAGATATACCAAG	569
Db	201962	GTCGAAGAACACGCGGCTGCGCTCGGTGATCCAGCGGATCGACCGGCTTCTCGAAGG	201903
QY	570	CTTTTGAGGCATTTTGTATCTTAACCTGACCTACAGATTTGGCTGAGCATCTTTTCT	629
Db	201902	CGCTGGGACCATGTCGATTCGAAAGTTCAGTCTGATGCGACACCGTGTGCTGTCT	201843
QY	630	TCTAGATTAAGGCGCTATTAAAGGTTTCTTGTGACATTTAGATTTTCTCTT	689
Db	201842	ATCTCGGACACCGGCGCGGTGAAGGCTTTCACATACCTACGCAATGGATCTCTGACA	201783
QY	690	CAATGTTTACGCGCTTTTCATGATCAATATTTTCTTCATGCTGTGATGAATTAAGACC	749
Db	201782	CGGTTTCAACCGCTTACCTTACCTGACGCGCTGCTGTGATGATGCTGCGCGCC	201723
QY	750	AACATACACAGTTGCATATG	769
Db	201722	GTCGGAAGGAATGCGCAAG	201703

RESULT 30

AX413017
LOCUS AX413017 349980 bp DNA linear PAT 14-JUN-2002
DEFINITION Sequence 8 from Patent WO228891.
ACCESSION AX413017
VERSION AX413017.1 GI:21445475
KEYWORDS
SOURCE
ORGANISM Listeria innocua.
Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
REFERENCE
AUTHORS Glaser, P. and Kunst, F.
TITLE Listeria innocua, genome and applications
JOURNAL Patent: WO 0228891-A 8 11-APR-2002;
Pasteur Institut (FR)
FEATURES
source
1. .349980
/organism="Listeria innocua"
/db_xref="taxon:1642"
/note="original seq 8 (495269) splitted in 2-new seq 8: 1
to 349980-seq 4026: old seq 8 from 300000 to 495269"
BASE COUNT 117290 a 59265 c 69878 g 103545 t 2 others
ORIGIN
Query Match 5.6%; Score 103.6; DB 6; Length 349980;
Best Local Similarity 48.8%; Pred. No. 6.5e-13;
Matches 280; Conservative 0; Mismatches 294; Indels 0; Gaps 0;
Qy 150 TCTCAGGAAATTTACCCACCGTGAAGTGAGCAAACTCGCCCTCAGATTAAATCTGGAG 209
Db 190895 TTTCGGGAAGCTTTACACTGAGAGCAAAAGATTAGCTGAATTTGTAATCTGGCG 190954
Qy 210 CGATGCTTTTGGTCCCGAGGTTCTCAGTGAAGACGATCTCTTCGATCTTGGGAAA 269
Db 190955 CACTTCCTGTAATAAAGTACTGAAGTTTACTCTACATCTGTTGGCGACAGATTGGTCAAG 191014
Qy 270 AACATGTACAAAGGCATTATCTCAGCATGCTGTGGCTTGCAATGCTATTGTTTGA 329
Db 191015 ATGCATTAACAAGAACTATTTAGCTGGTATTAATCGCGGTGTAAGCTATCTTTATTTTCA 191074
Qy 330 TGAGCGTATATTATAGATTGGAGGCGTCATCGCTTCGGGAGCTGTTCTCTGAATCTTT 389
Db 191075 TGATGCTGTTTACCGTCTACCGGCTGTTATCGCATCTATTACATAGTTGCTTACACTT 191134
Qy 390 TGCCTATCTGGCAGCTCTACAGTATTTGGATGCGCCACTACCTTGTGAGCAACTCC 449
Db 191135 APTTAGTTCTATTAATACTAAGCTTACTCAATGCCACACTTACTTTACCAGGATTGCG 191194
Qy 450 GGATTTCTTCTGCTATGGGATGGCGGTAGATGCAATGTTCTTGTATTTCGAAAGATCC 509
Db 191195 GCTTGATTTCTGGGTATAGGTATGGCTGTGACGCAACGTTTATAACATACGAGCGAATAA 191254
Qy 510 GAGAGGAATTTTATTTGCTCAAGTCTTAAATAATCTGTAGAAAAGGATATACCAAGG 569
Db 191255 AAGAAGAAATCAAGTTCGGAAGTTCGACAAAGACGACTTTCGAGTCGGTGGAAAAGAG 191314
Qy 570 CTTTGGAGCCATTTTGAATCTTAATCTGACTACAGTATGGCCCTCAGCACTCTTTTCT 629
Db 191315 CATTCGTCGCAATTTAGATGTAATTAACGACACTGATTGTCGAGTTCTTTCT 191374
Qy 630 TCTAGATACAGGCGCTATTAAAGGGTTGCTTCACATGATGATTTAGGAATTTCTCTT 689
Db 191375 ACTTCGGTACAGCTCTATCAAGGGTTTGGTACAGTTCTAATCATTAGTATTCTAGTCA 191434
Qy 690 CAATGTTTACGGCTCTTTTCATGACTAAATTTT 723
Db 191435 GCTTCTTAACAGCTGTTGGGTTCAAGATTTCTT 191468
RESULT 31
AX417045/c
LOCUS AX417045 349980 bp DNA linear PAT 14-JUN-2002
DEFINITION Sequence 4036 from Patent WO228891.
ACCESSION AX417045

AX417045.1 GI:21449655
LOCUS AX417045.1 349980 bp DNA linear PAT 14-JUN-2002
DEFINITION Sequence 4036 from Patent WO228891.
ACCESSION AX417045
VERSION AX417045.1 GI:21449655
KEYWORDS
SOURCE
ORGANISM Listeria innocua.
Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
REFERENCE
AUTHORS Glaser, P. and Kunst, F.
TITLE Listeria innocua, genome and applications
JOURNAL Patent: WO 0228891-A 4036 11-APR-2002;
Pasteur Institut (FR)
FEATURES
source
1. .349980
/organism="Listeria innocua"
/db_xref="taxon:1642"
/note="seq 2058, original length: 3.011.208 replaced
by-seq 2058: 0.000.001 to 0.349.980-seq 4032: 0.300.001 to
0.649.980-seq 4033: 0.600.001 to 0.949.980-seq 4034:
0.900.001 to 1.249.980-seq 4035: 1.200.001 to
1.549.980-seq 4036: 1.500.001 to 1.849.980-seq 4037:
1.800.001 to 2.149.980-seq 4038: 2.100.001 to
2.449.980-seq 4039: 2.400.001 to 2.749.980-seq 4040:
2.700.001 to 3.049.980-seq 4041: 3.000.001 to 3.011.208"
BASE COUNT 101566 a 72212 c 58560 g 117642 t
ORIGIN
Query Match 5.6%; Score 103.6; DB 6; Length 349980;
Best Local Similarity 48.8%; Pred. No. 6.5e-13;
Matches 280; Conservative 0; Mismatches 294; Indels 0; Gaps 0;
Qy 150 TCTCAGGAAATTTACCCACCGTGAAGTGAGCAAACTCGCCCTCAGATTAAATCTGGAG 209
Db 68060 TTTCGGGAAGCTTTACAACTGAAGAAGCAAAAGATTAGCTGAATTTGTAATCTGGCG 68001
Qy 210 CGATGCTTTTGGTCCCGAGGTTCTCAGTGAAGACGATCTCTTCGATCTTGGGAAA 269
Db 68000 CACTTCCTGTAATAAAGTACTGAAGTTTACTCTACATCTGTTGGCGACAGATTGGTCAAG 67941
Qy 270 AACATGTACAAAGGCATTATCTCAGCATGCTGTGGCTTGCAATGCTATTGTTTGA 329
Db 67940 ATGCATTAACAAGAACTATTTAGCTGGTATTAATCGCGGTATGCTATCTTTATTTTCA 67881
Qy 330 TGAGGCTATATTAGATTGGAGGCGTCATCGCTTCGGGAGCTGTTCTTCTGAATCTTT 389
Db 67880 TGATGCTGTTTACCGTCTACCGGCTGTTATCGCATCTATTACATAGTTGCTTTACACTT 67821
Qy 390 TGCCTATCTGGCAGCTCTACAGTATTTGGATGCGCCACTCACCTTGTGAGCACTCGCTG 449
Db 67820 ATTAGTTCTATTATTAATTAAGCTTACTCAATGCCACACTTACTTTACCAGGATTGCG 67761
Qy 450 GGATTTCTTCTGCTATGGGATGGCCGTAGATGCAAAATGTTCTTGTATTTCGAAAGAAATCC 509
Db 67760 GCTTGATTTCTGGGTATAGGTATGCTGTGACGCAACGTTTATAACATACGAGCGAATAA 67701
Qy 510 GAGAGGAATTTTATTTGCTCAAGTCTTAAATAATCTGTAGAAAAGGATATACCAAGG 569
Db 67700 AAGAAGAAATCAAGTTCGGAAGTTCGACAAAGCAGCACTTTGAAGTCGGTGGAAAAGAG 67641
Qy 570 CTTTGGAGCCATTTTGAATCTTAATCTGACTACAGTATTCGCGCTCAGCACTCTTTTCT 629
Db 67640 CATTCGTCGCAATTTAGATGTAATTAACGACACTGATTGTCGAGTTCTTTCT 67581
Qy 630 TCCTAGATACAGGCGCTATTAAAGGGTTTGGTTCGACATGATTTTAGGAATTTCTCTT 689
Db 67580 ACTTCGGTACAGCTCTATCAAGGGTTTGGTACAGTTCTAATCATTAGTATTCTAGTCA 67521
Qy 690 CAATGTTTACGGCTCTTTTCATGACTAAATTTT 723
Db 67520 GCTTCTTAACAGCTGTTGGGTTCAAGATTTCTT 67487
RESULT 32
AE006057/c
LOCUS AE006057 12201 bp DNA linear BCT 08-MAR-2001

EVTGGDTVFRLESAKKATDTRMTQOGLIYRRVDAAGTREPTEIMRGDRLIE
 VRIQSAOELKDICTTAKTTRHPVSTTSNPAAPASNEILLPDEROGIYLLDEY
 PVVIGDLDLIDARPTDNGAPAVSFREYNAGARADYRAGHIGEPFAYLDKXISA
 PTIQHAIACGSGIITGRFSIEATDALLIRAGALPMGTFRIBERTIGELGDSVKA
 GMAVAVIGVAVAVIAMIASYVALFNINIFIAVMAIGITMLPDIQIIV
 LRTGTVDNVLIERMERREIRSGSPVAIEGPKMSAIIIDAVNTSFLSAILFV
 LGAPVRGFVAVTMIGIAISIFAIWVRLMIVIMWGRPKTIV"

1897. .2598
 /gene="secf"
 1897. .2598
 /gene="secf"
 /note="Part of 'sec'-dependent transport complex in
 prokaryotes; found 5' to the helABOX genes required for
 cytochrome c biogenesis"
 /codon_start=1
 /transl_table=11
 /product="SecF"
 /protein_id="AB62802.1"
 /db_xref="GI:2252774"
 /translation="MAPRLKLVPEKTNFDFRQWATFGAAYMYMIAVYLPVIGLN
 FGIIDFGGTTIRRESTTALDVGVRALELELDIVTISEVRDPSFREDQHMAMTIO
 MOEDGOGAHOQOQOGLVAKETALVADPALKITSFESVGRVSGELVMTAVMSLL
 AATVIMFYIWRFEWQFALGAVVALVHDLTVGLFVQLFEDLTVAALITIGY
 SINDVVFIDRLRENLIKXTMPLRDVMVSETLSRTMTGTTLLALVPLWING
 DVIKGFVPMWGVFTGTVSYVAVKNIVLFIIDLNKKEKPSDKFNSGADGAP"

BASE COUNT 496 a 930 c 1015 g 559 t
 ORIGIN

Query Match 5.5%; Score 102.2; DB 1; Length 3000;
 Best Local Similarity 48.4%; Pred. No. 2.2e-12;
 Matches 284; Conservative 0; Mismatches 303; Indels 0; Gaps 0;

QY 150 TCTCAGGAAATTTACCAAGCTGAAGTGAACAACCTGCTCAATTTAAATCTGGAG 209
 DB 1262 TCACCGGCGCTTACAGATCAGAGACGAGACTGCGCCCTTTTGGCGCGGCG 1321
 QY 210 CGATGCTTTTGTCCGAGGTTCTCAGTGAAGAGAGATCTCTTCGATCTTGGGAAA 269
 DB 1322 CCCGCTCGCGGAGTACCTTCTCTTGAAGAACGACAGATGGCCGGAATCGGCGCG 1381
 QY 270 AACATGTACAGAGCATTTCTCAGCATGCTGTGGCTGGCAATCTTATTTTGA 329
 DB 1382 ACAGCGTCAAGCGGAGATGCTCCCTCGCATGCGCTTGTGGCGGTGCTGACCA 1441
 QY 330 TGAGCGATATTAATGATTTGAGAGCGCTCATGCTTGGGAGCTGTTCTTGAATCTTT 389
 DB 1442 TGATGCGCTCTCAGGCGCTTTTCCGCTTCTTCTCGTGGCGCTGTTCAACAATG 1501
 QY 390 TGCTTATCTGGGAGCTCTACAGATTTGATGGCCACACCTTGCAGAGATCGCGTG 449
 DB 1502 CTTTCATCTTCGCGGTATGAGCGCGAGTGGCGGACAGTACCTTGGCGGAGTCGCG 1561
 QY 450 GGATTTGTTCTGTATGGGATGAGCGGATGCAAAATGTTCTTGTATTCGAAAGATCC 509
 DB 1562 GGATGCTGCTACCATCGGACCTCGGTGATGCGAAGTCTGATCTAGAGAAAGCATGC 1621
 QY 510 GAGAAGATTTTATTTGTCTCAAAAGCTTTAAAAATCTGTAGAAAAAGATATACCAAG 569
 DB 1622 GCGAGGAATCCGGTGGGCAATTCGCGCGCGAGTGAAGTGGGCTTTGACAGG 1681
 QY 570 CTTTGGAGCATTTTGTATTTCTAATCTGACTACAGATTTGGCCTGAGCATTTCTTTCT 629
 DB 1682 CGATGCGGCGATATGCAJCGGAACGTGACCTGTCTCTCTCGCGCATCTGTCG 1741
 QY 630 TCCAGATACAGGCGCTATTAAGGTTTGTGTGACATGATTTTGAAGATTTTCTTCT 689
 DB 1742 TGTGCGGGCGGCGCGGTGCGCGCTTTGCGGTGACAGATGATCGGATCGCGCT 1801
 QY 690 CAATGTTTACGCTCTTTTATGACTAAATTTTCTTCATCTGTGG 736
 DB 1802 CGATCTTACCGCGATCTGAGTGGTGGCGGTGATGATGATCTGCG 1848

RESULT 34
 AX416793/c 5839 bp DNA linear PAT 14-JUN-2002
 LOCUS AX416793 Sequence 3784 from Patent WO0228891.
 DEFINITION AX416793
 ACCESSION AX416793
 VERSION AX416793.1 GI:21449250
 KEYWORDS
 SOURCE Listeria monocytogenes ATCC 19115.
 ORGANISM Listeria monocytogenes ATCC 19115.
 Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
 REFERENCE
 1 Glaeser, P. and Kunst, F.
 Listeria innocua, genome and applications
 Patent: WO 0228891-A 3784 11-APR-2002;
 JOURNAL Pasterur Institut (FR)
 FEATURES
 source location/Qualifiers
 1. 5839
 /organism="Listeria monocytogenes ATCC 19115"
 /db_xref="taxon:176281"

BASE COUNT 1757 a 1256 c 954 g 1872 t
 ORIGIN

Query Match 5.5%; Score 102.2; DB 6; Length 5839;
 Best Local Similarity 47.7%; Pred. No. 2.1e-12;
 Matches 299; Conservative 0; Mismatches 328; Indels 0; Gaps 0;

QY 117 TTTTAAACGTCCTCCATTGAAAAATCATGCCAGTGTCTCAGGAAATTTACCCACCGTGAAG 176
 DB 3730 TTATGTGCGGCTTGTATGATGATGAAATTTCCGGAATTTACACACGGAAGAG 3671
 QY 177 TGAGCAAACTCGGCTCAGATTTAAATCTGAGCAGATGCTTTGTGTTCCGAGTTCTCA 236
 DB 3670 CAAAAGTTAGCTGATTTGCTTAACTGTGTGCTACTCCCTGTTAAATGAAAGATT 3611
 QY 237 GTAGAGAGACATCTCTTGATCTTGGAAAAAACAATGATACACAGGCTTATCTCAG 296
 DB 3610 ACTGACATCTGTGTGCTCAATTTGGGCAAGATGCTTACAGAAACAAATTTTACGT 3551
 QY 297 CAGCTGTGCTTGGCAATGCTTATTTGTTGATGAGCGTATTTATGATTTGAGGCG 356
 DB 3550 GTATTTATGGGGTCATTTGCTTATTTTCAATGATGAGGATTCAGGTTACCGTTAC 3491
 QY 357 TCATGCTTGGGAGCGTCTCTGATCTTGTGCTTATCTGGGAGCTCTACAGATAT 416
 DB 3490 TTATGCTTGTATCATTATGTTGATGATCACTTATTTAGTACTTATATCTAGTCTGC 3431
 QY 417 TGATGCGCACCTCACCTTGTGAGACTGCTGGAGATGTTCTTGTGATGGGATGCGC 476
 DB 3430 TTAACGCAACACTTATCTTACAGGATTCAGGTTGATTTGGGGAATGATAGGCTG 3371
 QY 477 TAGATGAAATGTTCTTGTATTTGAAAGAAATCCGAGAGAAATTTTATGTCTCAAAGTC 536
 DB 3370 TTACGCGGAACGTAAATAACCTACAGAGCGATTAAGAAAGAAATCAAGTGGTAGTCA 3311
 QY 537 TTAATAAATCTGTAGAAAAAGATATACCAAGGCTTTGGAGCTTTTGTATTTACT 596
 DB 3310 CGAAAGCGCTTTTGAAGTGGCGGTAAAGAGATTTCCGTGAGATTTTGAAGTACT 3251
 QY 597 TGACTACAGTATTTGGCTCAGCACTTTTCTTCTTCTAGATACAGGCGCTTATTAAGG 656
 DB 3250 TGAGACACTTATCGTAGCTGCGGCTTCTTCTTCTTCTTGAACAGGCTCATCAAGG 3191
 QY 657 TTGCTTTGACATGATTTTGAAGAAATTTCTTCAATGTTTACGCGCTTTTATGACTA 716
 DB 3190 TTGTACAGGTTCTTATCATTAATTAATTTTACTGACCTTCTTGAAGCGGCTTGGGATCA 3131
 QY 717 AATTTTCTCATGCTGTGATGAATA 743
 DB 3130 GATTTCTTATAGGCTTATTTAGTAAAA 3104

RESULT 35
 ALS91979/c


```
/note="CT2261"
163..780
/genre="rnhb"
/note="identified by match to PFAM protein family HMM
PF01351"
/codon_start=1
/transl_table=1
/product="ribonuclease HII"
/protein_id="AAW73476.1"
/db_xref="GI:21648308"
/translacion="MLSTDEYPLWESLQVCGIDEAGRGPLAGPVAAAVVPRHPR
PTGIFAKLDSKLTALRDELAALAIRESAALGVVDAETDIRINILQATLMMNL
AVESLSTGPFLVDGNERFVLPVPIQTVKGSKFVSAASVLAKTHRELMTTS
AAVEPYGFEVHFYPARHVEAIARHGRCAIRHQSFLKRLGK"
800..1189
/genre="CT2262"
800..1189
/genre="CT2262"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="AAW73477.1"
/db_xref="GI:21648309"
/translacion="MNAQWLGAGEGKIAARHLAAAGYRIVARNYRHRNEIDIAFD
GEALCFEVTASLGHGAESVTRSKKEIAAPAAAGYLASLDDPWITCRFDVIAVL
ALSIDERSIRKYEIHEIKAAAFWVGDK"
1262..3205
/genre="gyzB"
/note="CT2263"
1262..3205
/genre="gyzB"
/note="identified by match to PFAM protein family HMM
PF00204"
/codon_start=1
/transl_table=1
/product="DNA gyrase subunit B"
/protein_id="AAW73478.1"
/db_xref="GI:21648310"
/translacion="MQRTDIQTQNASTEYCATNIQVLGIEHVRKRPAMVIGDIHSR
GLHLVVEIDNAIDETLAGVYVHMVNDGVTVDHGRGIPVDIHPVKKSALE
LMTVIGAGGKPKGAKVKGSLHVGCAVUNALSEWCEVYVDGKIPEQTVRRGVP
QGVVEITGTDQRTKVPFDPDEIFKITEFRKDIILDRMRELAFLNSNRIVQDAE
GNEIFHYEGGLKEFVRFIDNSRLSLKEPIFLSGERDSTMVEIALQINDSYENIFS
VYNNINHEGHTVGTFRKATRLNAYAKNDLLKNVITLTDGDFKEGLTAVISVK
VPEPFGQTKTLGNSGTQSVETIYNEQLAEFAESNPGTLKLIIEKVKSAISREA
ARAKELTRKSLGSEGLPKLADCSINDPDHCELIVVEGDSAGSAKQGRDRSFOA
ILPLKGLINVEARLHMLNENIEKITIILALCTSIGEEFSEPKLYRGLIIMTDA
VDGNHRTLLTTPFRMRSLIIBAGKVFIAPPLYLKSGREQQIADWEDELSIVEQ
MKKLQKGAWSIQRYKGLGEMNPEQLWSTMTDPAHRSLLQVTVENAEADQVFSTLM
GDKVEPRDRDFIEKNARYVRLDV"
complement(3194..4510)
/genre="surA"
/note="CT2264"
complement(3194..4510)
/genre="surA"
/note="identified by match to PFAM protein family HMM
PF00639"
/codon_start=1
/transl_table=1
/product="peptidyl-prolyl cis-trans isomerase SurA"
/protein_id="AAW73479.1"
/db_xref="GI:21648311"
/translacion="MKKVLFAVLAALMIAMGFADAASTGLDRIVAIVGNELIASD
VNEQELMLHQYETKDPOLRKRLLENMINOKIILTAKIDTVKVDKSDQDAAR
YSSIRAGFPVSAMESRPNVRLKQHIREDIRDQOMIEAFRKQFHEVTSIDETM
AFYDQKALPEAPETVSVSQIIMPLVSEARQAALDKIKAVQQLEAGGSFATLAR
EYSDPGSRGKGLDVGKRGELVSPFEAAVSLKPGQISGIVETFRGHIILIDKE
GLIHTHRIALDRSKTDIPATIALKLSIKDVLGSKATFAEMAKYSDDPASATNG
DRITSGSNPDLEVRPLRKIIDGLSKSGKDISQPEKIESDKSAPFYALFNLNSRT
PAHLVTEHDFAOIELANLHKSQELFNANIEKLKKEVVVWMSDI"
complement(4769..7544)
/genre="CT2265"
```

/note="this region contains an authentic frame shift and is not the result of a sequencing artifact; similar to GB:AB000513; identified by sequence similarity; putative; protein-export membrane protein secD, putative"

BASE COUNT	1761 a	2213 c	2162 g	1551 t
ORIGIN				

Query Match	5.4%	Score 100.6;	DB 1;	Length 7687;
Best Local Similarity	51.7%	Pred. No. 4.5e-12;		
Matches 229;	Conservative	0;	Mismatches 214;	Indels 0;
Gaps	0;			

QY	301	CTGTGGCTTGCAGTCTATTGTTTTCATCAGCGGTATATATATAGATTTCAGGCGGTCAAT	360
DB	6186	CTGGGGTTTACGCGCGTTCGCTATCTTCATGATTGTTTATTATAACAGGCGGTATCGC	6127
QY	361	CGCTTCGGGAGCTCTTCTTCTGAATCTTTTCTGCTATCTGGGCAGCTCTACAGTATTTGGA	420
DB	6126	CGCCGACCTGGCGCTCGTCTCAACATTATGGTCATCATCGCGTGTCTGCGGGTTTCAG	6067
QY	421	TGCGCCACTACCTTGTTCAGACTCGCTGGGATTGTTCTTGTCTATGGGATGCGCGGTAGA	480
DB	6066	CGCCGTACTGACCTTACCCGGTATTGCGGTATCGTCTGACGATGGTATGGCTGTGA	6007
QY	481	TGCAATGTTCTTGTATTTCGAAAGAAATCCGAGAGGAATTTTATTGTCTCAAAGTCTTAA	540
DB	6006	CGCCAACGTGCTGATTACGAGCGCGTCCGCGAGGAGCTTTCGCGACGCGCAAGAGCTCG	5947
QY	541	AAATCTGTAGAAAAAGATATACCAAGGCTTTTGGAGCCATTTTGTATTCTAACTTTCAC	600
DB	5946	CATGGCTATCGACGCGGGGTACACCGTGGTGTTCATCGATTCGACTCGCAGCTCAC	5887
QY	601	TACAGTATGGCTTCAGACCTTTCTTTCTTCTAGATACAGGCGCTATTAAAGGTTTGC	660
DB	5886	GACGCTTGGCGCAGCTTTCTGCTCTATATTACGCGGTGGGCGCAATTCAGGATTTCG	5827
QY	661	TTTCACATTGATTAGCAATTTCTTCTCAATGTTTACGCTCTTTTCATGACTAAATT	720
DB	5826	CGTCACCTCATGATCGGTACTGCGCGAGCTTGTTTACGCGGATCGTCGTCAACAGG	5767
QY	721	TTTCTTCATGCTGGGATGAATA	743
DB	5766	GCTGTTGACCTGCTGCTCGACA	5744

RESULT 37	AE004160	14261 bp	DNA	linear	BCT 31-JUL-2000
LOCUS	AE004160	Vibrio cholerae chromosome I, section 68 of 251 of the complete chromosome.			
DEFINITION	AE004160 AE003852				
ACCESSION	AE004160.1	GI:9655181			
VERSION					
KEYWORDS		Vibrio cholerae.			
SOURCE		Vibrio cholerae.			
ORGANISM		Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.			
REFERENCE		Heidelberg, J.F., Eisen, J.A., Nelson, W.C., Clayton, R.A., Gwinn, M.L., Dodson, R.J., Haft, D.H., Hickey, E.K., Peterson, J.D., Umayam, L.A., Gill, S.R., Nelson, K.E., Read, T.D., Tettelin, H., Richardson, D., Ermolaeva, M.D., Vamathevan, J., Bass, S., Qin, H., Dragoi, I., Sellers, P., McDonald, L., Utterback, T., Fleischmann, R.D., Nierman, W.C. and White, O.			
AUTHORS		DNA sequence of both chromosomes of the cholera pathogen Vibrio cholerae			

TITLE	Nature 406 (6795), 477-483 (2000)
JOURNAL	20406833
MEDLINE	10952301
PUBMED	2 (bases 1 to 14261)
REFERENCE	Heidelberg, J.F., Eisen, J.A., Nelson, W.C., Clayton, R.A., Gwinn, M.L., Dodson, R.J., Haft, D.H., Hickey, E.K., Peterson, J.D., Umayam, L.A., Gill, S.R., Nelson, K.E., Read, T.D., Tettelin, H., Richardson, D., Ermolaeva, M.D., Vamathevan, J., Bass, S., Qin, H., Dragoi, I., Sellers, P., McDonald, L., Utterback, T., Fleischmann, R.D.,
AUTHORS	

TITLE
JOURNAL

Nierman, W. C., White, O., Salzberg, S. L., Smith, H. O., Colwell, R. R.,
Mekalanos, J. J., Venter, J. C. and Fraser, C. M.
Direct Submission
Submitted (14-JUN-2000) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA

FEATURES
Location/Qualifiers

1..14261
/organism="Vibrio cholerae"
/strain="H16961"
/serotype="O1"
/db_xref="taxon:666"
/chromosome="I"
/note="biotype: El Tor"
complement(107..616)
/gene="VC0737"
complement(107..616)
/note="similar to GB:LI7309 SP:P39066 PID:348051
PID:2293317 GB:AL009126; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="acetoin utilization protein Acub, putative"
/protein_id="AAF93902.1"
/db_xref="GI:9655182"
/translation="MSGKVRDPGSDGSLRKKELAMTKVEDMMTRHPHTLRTTLNDA
KHLMEALDIRHPIVDANKKLISGSRDLAAQESLSQASQDSLAPEPTPIPEYPM
TDVTSVAPCAGLKEASAIWMQKHIGCLPVAKQVLGIIIDSDFTTAINLLEQES
EPDELDEEC"
860..1904
/gene="VC0738"
860..784
/gene="VC0738"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF93903.1"
/db_xref="GI:9655183"
/translation="MRTVESVGEPDGLQKVTLPEVYHHAFLYLENT"
846..1904
/gene="VC0739"
846..1904
/gene="VC0739"
/note="similar to GB:M37702 SP:P21516 PID:147485 GB:U00096
PID:1773089; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="S-adenosylmethionine:tRNA
ribosyltransferase-isomerase"
/protein_id="AAF93904.1"
/db_xref="GI:9655184"
/translation="MQVSDFFHLLPDELLIARYPPAERTASRLQLDQNSQGLVDGTEK
DVLIELVPDILLVFNTRVDPARMGRKASGSKLEVERMLDEHTILLAVRSKPEK
PGTELYGENDERANWQARHDLPELRFAETVVLIDINOIGMPLPYIDRDEEA
DKERYQTVNCKPGAVNAPTRGHPDQALQIOAKGYELAYVTLHAGAGPOGVARE
NIHDHMAEVEVPOEVDAITATKARGKVAVGTVSASLSAADALQKTEIK
PFPGDTEIFIPPGYOYOLVDCLITNPHLPSTLIMLVSAFAGYEHTMAAYEHAIVEQY
RPFSGDAMFIRKOMTKA"
complement(1859..2122)
/gene="VC0740"
complement(1859..2122)
/gene="VC0740"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF93905.1"
/db_xref="GI:9655185"
/translation="MSEKSDHFYPPNRANLRSQOVNSSLIAINKALCGAVAPCEK
GLHLFMILPALVFIPICHLKQPLGVIRIDVALVLCRLINIASP"
2133..3272
/gene="VC0741"

CDS

2133..3272
/gene="VC0741"
/note="similar to PID:581229 GB:U00096 PID:1773090
PID:1786607; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="guanine tRNA-ribosyltransferase"
/protein_id="AAF93906.1"
/db_xref="GI:9655186"
/translation="MKLKEELKKKGNARRGOLIPERGTVQPAFMPVGYGVKGMT
DEVERENAOILLGNTPHLMIRPOGVWKKMGKDLDFNMWGGPILIDSGGQVPSLD
IRKTRBEGVFRNPNNGDKIRMDAKSEKIQKDSLDIYMLFDECTPYRPHDEAKKS
MEMSLRWAKRSRSDHDKLEPNVNLFGIQQGVYEDLRVSKGLTEIPDGAVCGLA
VGEPEKEDHVRVLEHTCPQLPEDKERYLMGVGPELDVGVGRGIDMPCCVETNARN
GHLFVYGVGVIKIRNAHKTDTPLDPHCDYCKNYSXYLHLDRCNCEIIGARLNTI
NHLRYORLMESIRKAIDEDRDPQVFVAEFARARNRNPPLQDKA"
3377..3709
/gene="VC0742"
3377..3709
/gene="VC0742"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF93907.1"
/db_xref="GI:9655187"
/translation="MSLISVAHAGEGAPQGGFEMIIMLAFVIFYFMIYRPOAKR
VKEHNMAAMAKGDEVLTSGLVGRITKIAEDNAYITELTNNNEVVIKKQFVAVL
PKGTLKSL"
3731..5584
/gene="VC0743"
3731..5584
/gene="VC0743"
/note="similar to SP:P19673 GB:S68715 GB:X56175 PID:581230
PID:545175; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="protein-export membrane protein SecD"
/protein_id="AAF93908.1"
/db_xref="GI:9655188"
/translation="MKNRYPIMKYIMVMVLTIVAAVLALPNYIGEDPAIQTGARGAS
VMSITDAVTDLANKAQLSQKSIALENGSIIVRNDTQVQISARDISFALGDKIVA
LNLAPSTPYMLSESIGAPMLKGLIDLGKGVHFLMEVDNDAEMLYSQQEAPRSIDRD
EKIRYAIRPLSDAVEVTLRDAEQALVTLLESKRDMTFTTSDEGFVLVAKTE
ARLOEIRNVAECONITILRNRYNELGVAEPLVQROGATRIIVELPGVOTARAKETLG
ATATLEFRVDDKADLAAAGAPAGSGSEIKFDRNGRPVLLKKRVILGSSSIDASS
ADBYGRQVNISIDSEGNKMGAFSKYKQIMATYAFAYKSGKSPGKVLITGE
EVINQATIOSALGRNFRITGIDSPAAHNLALLRAGALIAPISTYBERTIGSMQO
NIDMGIOACIGWGVAVMLFTVLYRKFEMIANTALMANVLIIGVSMIPGATMTLPG
IAGIVLTGVAVDANVLIIFERIRELRBEKNPQALHQGVANAFSTIADANITTLITTA
ILFAVGTGAIKGFVTLISGILTSMTFAIVGRICVNLVGGKRIKLSI"
5599..6546
/gene="VC0744"
5599..6546
/gene="VC0744"
/note="similar to SP:P19674 GB:X56175 PID:581231 GB:U00096
PID:1773093; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="protein-export membrane protein SecE"
/protein_id="AAF93909.1"
/db_xref="GI:9655189"
/translation="MFOILKADKIDFMRSKFAFALSLVMAISIFTLSTKMLNGL
DPFGCLTIEVGFEPQENLEQIRBSAGKGDVTONFGSAPDVMTLRGRDMAGGA
RLAQVILAIKEGTGQNVEMKRIEFGVNPQDDELTEAGALIAVSLICILLVSVREFM
RLAAGVIALAHADVITTLGIFSLQIEVDLITVALLTVGYSLNITIVFPDIRREMF
RKMRKEPEAIINMSITQTLSTRLLTSGLTFVVALFTVGGAMHIGFALALLGLITV
GYSSIVASALAKLIGITREHLLPPOVEKEGELDEMP"
complement(6714..7580)
/gene="VC0745"
complement(6714..7580)
/gene="VC0745"
/note="similar to GB:M34828 SP:P22783 PID:147899 GB:U00096

Db 2625 TTCATATCAGGTATTTACTCTACCTGGATTGGCGGATTGTTTGAAGTATG 2566
 QY 473 GCCGTAGATGCAATGTTCTTGATTCGAAAGATCCAGAGGAATTTTATGCTCA 532
 Db 2665 GCTGTGATGCAATATCATATGATATGATTAAGATGAACATGAATGAGC 2506
 QY 533 AGCTTAAAAATCTGTGAAAAAGATATACCAAGCTTTTGAGGCAATTTGATTC 592
 Db 2605 AGCTTAAACAGGTATTCAGAACATTAAGTTATTTACTATATTTGATTC 2446
 QY 593 AACTGACTAGATTTGGCTCAGACACTTTCTTCTTCTAGATACAGGCTATTTAA 652
 Db 2445 AACTTAACAACGTGATCGCGAGCTGTGCTTTCTTTTGAGAAAGTTCACTCA 2386
 QY 653 GGGTTGCTTGAACATTTAGGAATTTCTTCAATGTTTACGCTTTTCAATG 712
 Db 2385 GGGTTCGCAACCATGTTACTCTTAGGTATTTTAAATGATTTGTAAACGCAATTTCTTA 2326
 QY 713 ACTAATTTTCTT 726
 Db 2325 TCAAGAGGTTGTT 2312

RESULT 41
 AX144739/c 3511 bp DNA linear PAT 31-MAY-2001
 LOCUS AX144739 Sequence 3461 from Patent WO0134809.
 DEFINITION AX144739
 ACCESSION AX144739
 VERSION AX144739.1 GI:14283304
 KEYWORDS
 SOURCE synthetic construct.
 ORGANISM synthetic construct.
 REFERENCE 1 (bases 1 to 3511)
 AUTHORS Kimmerly, W.J.
 TITLE Staphylococcus epidermidis nucleic acids and proteins
 JOURNAL Patent: WO 0134809-A 3461 17-MAY-2001;
 GLAXO GROUP LIMITED (GB)
 FEATURES
 source Location/Qualifiers
 1..3511
 /organism="synthetic construct"
 /db_xref="taxon:32630"
 /note="synthetic nucleic acid sequence"
 BASE COUNT 1155 a 610 c 483 g 1263 t
 ORIGIN

Query Match 5.3%; Score 99.6; DB 6; Length 3511;
 Best Local Similarity 48.7%; Pred. No. 8.2e-12;
 Matches 270; Conservative 0; Mismatches 284; Indels 0; Gaps 0;

QY 173 GAAGTGAACAACCTGCCTAGATTAAATCTGGAGCGATGCTTTGTTCCGAGTT 232
 Db 2865 GAAGGAAACAAATGAGTATTAATGCGGCTCATTAACGATTGATTTAAAGAA 2806
 QY 233 CTCAGTGAAGAGCATCTCTTCTGATCTGGGAAAAAACAATGACACAGCATATC 292
 Db 2805 ATTACTCTACTCTGTGGTGACAAATTTGTCAGAGTCTTGATTAAGCAATGTTT 2746
 QY 293 TCAGATCTGTGGCTTGGCAATGCTTATTTGATGAGCGATATTAATAGATTGGA 352
 Db 2745 GCATCAATTTAGATTAATGATTAATTTATTTATGCTGTGCTTATGCTTGGCT 2686
 QY 353 GGGCTATGCTTGGGAGCTGTTCTTGAATCTTTTGTGCTATCTGGGACGCTTACAG 412
 Db 2685 GGTTAGTGAATCATGCTTAAACCACTTATATTATTTAACTTTAGTGCATTCAT 2626
 QY 413 TATTGATGCGCCACTCACTTCGACAGACGCTGGAGATTGCTTGTGATGGGGATG 472
 Db 2625 TTCAATCAGGTATTAATCTTACTGATGCGGAGATTAAGTTTAAAGTGAATG 2566
 QY 473 GCGGTAGATGCAATGTTCTTGATTCGAAAGATCCAGAGGAATTTTATGCTCA 532
 Db 2665 GCTGTGATGCAATATCATATGATATGATTAAGATGAACATGAATGAGC 2506

QY 533 AGCTTAAAAATCTGTGAAAAAGATATACCAAGCTTTTGAGGCAATTTGATTC 592
 Db 2605 AGCTTAAACAGGTATTCAGAACATTAAGTTATTTACTATATTTGATTC 2446
 QY 593 AACTGACTAGATTTGGCTCAGACACTTTCTTCTTCTAGATACAGGCTATTTAA 652
 Db 2445 AACTTAACAACGTGATCGCGAGCTGTGCTTTCTTTTGAGAAAGTTCACTCA 2386
 QY 653 GGGTTGCTTGAACATTTAGGAATTTCTTCAATGTTTACGCTTTTCAATG 712
 Db 2385 GGGTTCGCAACCATGTTACTCTTAGGTATTTTAAATGATTTGTAAACGCAATTTCTTA 2326
 QY 713 ACTAATTTTCTT 726
 Db 2325 TCAAGAGGTTGTT 2312

RESULT 42
 AF269550/c 3705 bp DNA linear BCT 01-AUG-2000
 LOCUS AF269550 Staphylococcus epidermidis strain SRI clone step.1010b02 genomic
 DEFINITION AF269550
 ACCESSION AF269550
 VERSION AF269550.1 GI:9623444
 KEYWORDS
 SOURCE Staphylococcus epidermidis.
 ORGANISM Staphylococcus epidermidis.
 REFERENCE 1 (bases 1 to 3705)
 AUTHORS Kimmerly, W.J., Taylor, J. David., Nelson, A.J., Godlevski, M.M.,
 Rubino, M.A., Nelson, F.J., Rivers, P.R., Tornella-Miller, I.,
 Listendee, S., Ashanti, C., Altschuller, G., Mam, L., Shepherd, N.S.,
 Fuchs, R., Fleming, T., Guan, X., Du, L., Cain, D.H., Miller, G.S. and
 Furdon, P.J.
 TITLE Transposon-mediated sequencing of the Staphylococcus epidermidis
 genome
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 3705)
 AUTHORS Taylor, J. David., Kimmerly, W.J., Nelson, A.J., Godlevski, M.M.,
 Rubino, M.A., Nelson, F.J., Rivers, P.R., Tornella-Miller, I.,
 Listendee, S., Ashanti, C., Altschuller, G., Mam, L., Shepherd, N.S.,
 Fuchs, R., Fleming, T., Guan, X., Du, L., Cain, D.H., Miller, G.S. and
 Furdon, P.J.
 TITLE Direct Submission
 JOURNAL Submitted (22-MAY-2000) Departments of Genomic Sciences and
 Bioinformatics, Genetics Directorate, Glaxo Wellcome, Inc., 5 Moore
 Drive, Research Triangle Park, North Carolina 27709-3398, USA
 FEATURES
 source Location/Qualifiers
 1..3705
 /organism="Staphylococcus epidermidis"
 /strain="SRI"
 /db_xref="taxon:1282"
 /clone="step.1010b02"
 BASE COUNT 1234 a 621 c 509 g 1341 t
 ORIGIN

Query Match 5.3%; Score 99.6; DB 1; Length 3705;
 Best Local Similarity 48.7%; Pred. No. 8.2e-12;
 Matches 270; Conservative 0; Mismatches 284; Indels 0; Gaps 0;

QY 173 GAAGTGAACAACCTGCCTAGATTAAATCTGGAGCGATGCTTTGTTCCGAGTT 232
 Db 2142 GAAGGAAACAAATGAGTATTAATGCGGCTCATTAACGATTGATTTAAAGAA 2083
 QY 233 CTCAGTGAAGAGCATCTCTTCTGATCTGGGAAAAAACAATGACACAGCATATC 292
 Db 2082 ATTACTCTACTCTGTGGTGACAAATTTGTCAGAGTCTTGATTAAGCAATGTTT 2023
 QY 293 TCAGATCTGTGGCTTGGCAATGCTTATTTGATGAGCGATATTAATAGATTGGA 352
 Db 2022 GCATCAATTTAGATTAATTTATTTATTTATTTATTTGCTGTGCTTCTTCTGCT 1963


```
QY 353 GCGTCATCGCTCGGGAGCTGTTCTTCTGAATCTTTTGCTTATCTGGCAGCTCTACAG 412
Db 1962 GGTTAGTGCAATCATGTCCTTAACCACTATATATTTAATTTAATCTTAGTCGCATCAAT 1903
QY 413 TATTGGATGCGCCACTCACCTGTCAGGACTCGCTGGGATGTTCTTGTCTATGGGATG 472
Db 1902 TTCAATACAGGTGATTAACTCTACCTCGATTGGCGCATTTAGTTAGGTAGGTATG 1843
QY 473 GCGTAGATGCAATGTTCTTGATTCGAAAGATCCGAGAGGAATTTTATATGCTCTCAA 532
Db 1842 GCTGTCGATGCCAATATCATATGATGATGAACGTTATTAAGATGAACCTAAGAAATGGACGC 1783
QY 533 AGTCTTAAAAAATCTGTAGAAAAAGATATACCAAGGCTTTTGGAGCCATTTTGTATCT 592
Db 1782 ACCTTAAACCAAGCGTATTCAAAAGCAAAATAAAAGTTCAATTTAACTATATTTGATTC 1723
QY 593 AACTGACTACAGTATTGGCTCGACACTTCTTTCTTCTTCTAGATACAGGCGCTATTAAA 652
Db 1722 ACCTTAAACCAAGCGTATTCAAAAGCAAAATAAAAGTTCAATTTAACTATATTTGATTC 1723
QY 593 AACTGACTACAGTATTGGCTCGACACTTCTTTCTTCTTCTAGATACAGGCGCTATTAAA 652
Db 1722 ACCTTAAACCAAGCGTATTCAAAAGCAAAATAAAAGTTCAATTTAACTATATTTGATTC 1723
QY 653 GGGTTGCTTTCGATTTAGGAAATTTCTTCTTCAATGTTTACGGCTCTTTTCATG 712
Db 1662 GGGTTGCGCAACCATGTTACTCTTAGGTATTTAATGATATTTGTACCGCAGTATTCTTA 1603
QY 713 ACTAAATTTTCTT 726
Db 1602 TCAAGAGGGTTGTT 1589

RESULT 43
AX144870/c
LOCUS AX144870 3705 bp DNA linear PAT 31-MAY-2001
DEFINITION Sequence 3592 from Patent WO0134809.
ACCESSION AX144870
VERSION AX144870.1 GI:14283435
KEYWORDS synthetic construct.
SOURCE synthetic construct.
ORGANISM artificial construct.
REFERENCE 1 (bases 1 to 3705)
AUTHORS Kimmery, W.J.
TITLE Staphylococcus epidermidis nucleic acids and proteins
JOURNAL Patent: WO 0134809-A 3592 17-MAY-2001;
GLAXO GROUP LIMITED (GB)
FEATURES
    source
        1. .3705
            Location/Qualifiers
                /organism="synthetic construct"
                /db_xref="taxon:32630"
                /note="synthetic nucleic acid sequence"
BASE COUNT 1234 a 621 c 509 g 1341 t
ORIGIN
Query Match 5.3%; Score 99.6; DB 6; Length 3705;
Best Local Similarity 48.7%; Pred. No. 8.2e-12;
Matches 270; Conservative 0; Mismatches 284; Indels 0; Gaps 0;

QY 173 GAAGTGACAACTCGCTCAGATTTAAATCTGGAGCGATGCTTTTGTTCGCGAGTT 232
Db 2142 GAAGCGAAACAATAGCTGAGTTATTAATCCGCTCATTTACCGATGTTTAAAGAA 2093
QY 233 CTAGTGAGAGAGCATCTCTTCTGATCTTGGGAAAAACAATGATACAAAGGCAATTATC 292
Db 2082 ATTTACTCTAATCTGTTGGTGCAAAATTTGGTCAAGATGCTCTTGATAGACCATGTT 2023
QY 293 TCAGCATGCTGTGCTGGCAATGCTTATGTTTGTATGACGATATATTAGATTGGA 352
Db 2022 GCATCAATTTAGGTATAGCAATTAATTTATTTATGCTTGTGTTCTATCGTTGCT 1963
QY 353 GCGTCATCGCTCGGGAGCTGTTCTTCTGAATCTTTTGCTTATCTGGCAGCTCTACAG 412
Db 1962 GGTTAGTGCAATCATGTCCTTAACCACTATATATTTAATTTAATCTTAGTCGCATCAAT 1903
QY 413 TATTGGATGCGCCACTCACCTGTCAGGACTCGCTGGGATGTTCTTGTCTATGGGATG 472
```

```
Db 1902 TTCAATACAGGTGATTAACTCTACCTGGATTGGCGCATTTAGGTAGGTATG 1843
QY 473 GCGTAGATGCAATGTTCTTGATTCGAAAGAAATCCGAGAGGAATTTTATGCTCTCAA 532
Db 1842 GCTGTCGATGCCAATATCATATGATGATGAACGTTATTAAGATGAACCTAAGAAATGGACGC 1783
QY 533 AGTCTTAAAAAATCTGTAGAAAAAGATATACCAAGGCTTTTGGAGCCATTTTGTATCT 592
Db 1782 ACCTTAAACCAAGCGTATTCAAAAGCAAAATAAAAGTTCAATTTAACTATATTTGATTC 1723
QY 593 AACTGACTACAGTATTGGCTCGACACTTCTTTCTTCTTCTAGATACAGGCGCTATTAAA 652
Db 1722 ACCTTAAACCAAGCGTATTCAAAAGCAAAATAAAAGTTCAATTTAACTATATTTGATTC 1663
QY 653 GGGTTGCTTTCGATTTAGGAAATTTCTTCTTCAATGTTTACGGCTCTTTTCATG 712
Db 1662 GGGTTGCGCAACCATGTTACTCTTAGGTATTTAATGATATTTGTACCGCAGTATTCTTA 1603
QY 713 ACTAAATTTTCTT 726
Db 1602 TCAAGAGGGTTGTT 1589

RESULT 44
AE013081
LOCUS AE013081 4357 bp DNA linear BCT 09-MAY-2002
DEFINITION Thermoanaerobacter tengcongensis strain MB4T, section 108 of 244 of the complete genome.
ACCESSION AE013081 AE008691
VERSION AE013081.1 GI:20516185
KEYWORDS Thermoanaerobacter tengcongensis.
ORGANISM Thermoanaerobacter tengcongensis.
REFERENCE 1 (bases 1 to 4357)
AUTHORS Bao, Q., Tian, Y., Li, W., Xu, Z., Xuan, Z., Hu, S., Dong, W., Yang, J., Chen, Y., Xue, Y., Xu, Y., Lai, X., Huang, L., Dong, X., Ma, Y., Ling, L., Tan, H., Chen, R., Wang, J., Yu, J. and Yang, H.
TITLE A Complete Sequence of the T. tengcongensis Genome
JOURNAL Genome Res. 12 (5), 689-700 (2002)
MEDLINE 21992816
PUBMED 11997336
REFERENCE 2 (bases 1 to 4357)
AUTHORS Bao, Q., Xu, Z., Hu, S., Dong, W., Chen, Y., Wang, J., Yu, J. and Yang, H.
TITLE Direct Submission
JOURNAL Submitted (07-OCT-2001) Beijing Genomics Institute/Genomics and Bioinformatics Center, Institute of Genetics and Development, Chinese Academy of Sciences, Beijing Airport Industrial Zone B6, Beijing 101300, China
REFERENCE 3 (bases 1 to 4357)
AUTHORS Li, W., Xuan, Z., Yang, J., Ling, L. and Chen, R.
TITLE Direct Submission
JOURNAL Submitted (07-OCT-2001) Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China
REFERENCE 4 (bases 1 to 4357)
AUTHORS Tian, Y., Xue, Y., Xu, Y., Lai, X., Huang, L., Dong, X., Ma, Y. and Tan, H.
TITLE Direct Submission
JOURNAL Submitted (07-OCT-2001) Institute of Microbiology, Chinese Academy of Sciences, Beijing 100080, China
FEATURES
    source
        1. .4357
            Location/Qualifiers
                /organism="Thermoanaerobacter tengcongensis"
                /strain="MB4T"
                /db_xref="taxon:119072"
                /note="putative rho-independent transcription terminator"
terminator 140..159
gene 174..1596
    /gene="AelB3"
    /note="TTE1186"
RBS 174..178
    /gene="AelB3"
```

```
CDS
187..1596
/gene="Aa1B3"
/Note="Best Blastp hit = gi|7462068|pir|H72266
aasB/chur-related protein - Thermotoga maritima (strain
MSB8) gi|4981883|gb|AAD36396.1|AE001787.1 (AE001787)
aasB/chur-related protein [Thermotoga maritima], score
99.8, E-value 7.00E-20"
/codon_start=1
/transl_table=1
/product="Arylsulfatase regulator (Fe-S oxidoreductase)"
/protein_id="AAM24417.1"
/db_xref="GI:20516186"
/translation="MSATMKFKRLGNIIVDPVSGAIHVVDVVDVVDLVYENHSRE
EIVNLKDKYKEEDILALISEVDLKGNGLLFTEDYKDLAISADSVIYAMCLNVAH
DCNLRCKYCPASTGENFKGRKMDFTGRKADIDFLISSGKRNTIEDPFGDELPAH
EYVKOLVEYGRKAKENKKNKFTITNAVALDDEKIEYENENSVNLSDGKEVN
DOMVRADGSGTYVIVPKIQPKAKGKKEYVGRFTAKNLDPEVDVLIADLGY
EISVEPVVEDKDDYTLKEEHLRIPEEYRLAEYIRYEEGRPFAPYFHKIDLKGG
PCTKRLQGGAGFEYTAIVPDEIRYCHQFVGEERKGLTDEGITNIELOKRWES
DIYREECANCMARFYCSGGCFANNININGINKPYKLACEMQRRILENAIAYITL
LRGKGDYCRVQRDKAAR"
/gene="Paay2"
/Note="Paay2"
/Note="TTE1187"
1540..1544
/gene="Paay2"
1553..2110
/gene="Paay2"
/Note="Best Blastp hit = gi|2493491|sp|Q57752|Y304_METUA
HYPOHETICAL PROTEIN Mj0304 gi|2127899|pir|A64338
ferriyochelin binding protein homolog - Methanococcus
jannaschii gi|1591027|gb|AA898291.1|U67485)
ferriyochelin binding protein (fbp) [Methanococcus
jannaschii], score 169, E-value 2.00E-41"
/codon_start=1
/transl_table=1
/product="Carbonic anhydrases/acetyltransferases,
isoleucine patch superfamily"
/protein_id="AAM24418.1"
/db_xref="GI:20516187"
/translation="MIIEYKGIKROIIDEAYIAETAEIIGDVEIKKNVIMGAVLR
GDVNRIVREGTINDCNVAVHDGHPCTGKCTTGAGAIYHAKCKGNVLTGGMGI
ILDDAEIGDNCTIGAKALVYGGKKIPGSGVLYGSPAVYVQLTBEELIESIHKSREHY
ELAKLHFSERQGLTVNKSNIENS"
/gene="SecD"
/Note="TTE1188"
2122..2126
/gene="SecD"
2134..3366
/gene="SecD"
/Note="Best Blastp hit = gi|13471172|ref|NP_102741.1|
probable protein-export membrane protein [Mesorhizobium
loti] gi|14021916|db|BAB48527.1| (AP002996) probable
protein-export membrane protein [Mesorhizobium loti],
score 259, E-value 4.00E-68"
/codon_start=1
/transl_table=1
/product="Preprotein translocase subunit SecD"
/protein_id="AAM24419.1"
/db_xref="GI:20516188"
/translation="WRSRGFIKPSVILVSLVAVYAPFGIKAGNFSISPIODHRIK
LDLGGVYVLEAQNVTODAIINKAIVIRNRVADLGVTOVIAQQOSNRIRVLPKM
KDPDKALEIIGKTAKLVGPDGVILITGANVSKSVYQSTGQOPVTLVLDAG
TKKFAEAKIQLGQPIALILDDVYSAVYVDVITGNNAVITGKQDPOESELATLR
AGSPVTLKRIAYSSGATGSPALKAISVANGICTLIWFMMAFYRLPGFTADLR
LITLINFIIYAINVTIDLPINGPLISIGMAVDITNIIIPERKEELMAGKTRPRL
DAGFAKMLRAVIDANVTIIIGAILIFYGSGNVGAFALITLLIGVSSSLFPAITVTRPL
LHALLEMLITRNVYFYGA"
/gene="SecF"
/Note="TTE1189"
3365..4290
/gene="SecF"
3365..3369
```

```
CDS
/gene="SecF"
3379..4290
/Note="Best Blastp hit = gi|12720452|gb|AAK02310.1|
(AK006057) SecF [Pasteurella multocida], score 160,
E-value 2.00E-38"
/codon_start=1
/transl_table=1
/product="Preprotein translocase subunit SecF"
/protein_id="AAM24420.1"
/db_xref="GI:20516189"
/translation="MNGHRFHIDVIGTKIMFTISGIMILIGIIMFVNGWIDPT
GGTIRMEFKIGKPEDTDKIKINDYKVRKYOIOXIKTGKDHVSIKTSPISDORLAI
IKREKKYNLTESDILMSOQVGSLSAERSGMILALVYSLWMLTYLGRPFEMSLA
AYIGLHNFTLISYVAFRITVDSFFIALTLVTSYSLHDTIVFPDRIRNDKWRMR
ADYAEVANSVQTLVRSINNVLTIVLIMLLVPLGPKALKOPALPLIGITWGAYS
IFSTPLWLINRREKRKASVYKAAK"
3457..4248
/gene="SecF"
/Note="Pfam match to entry SecD_SecF, Protein export
membrane protein, score 264.2, E-value 1.70E-75"
BASE COUNT 1406 a 644 c 1038 g 1269 t
ORIGIN
Query Match 5.3%; Score 99; DB 1; Length 4357;
Best Local Similarity 48.8%; Pred. No. 1.1e-11;
Matches 296; Conservative 0; Mismatches 310; Indels 1; Gaps 1;
QY 169 CCGTAGATAGCAAACTCCCTCAGATTAAATCTGAGAGCATGCTTTTCCCGA 228
Db 2754 CCAAGAGCCCTGGAATCTGCTACTTTGATTAGGAGCGGCTTTGCTCTCACTTAA 2813
QY 229 GGTTCAGAGAAAGACGATCTCTTCTGATCTTTGGAAAAAACAATGACAGGCAAT 288
Db 2814 GCGTATAGCCATATAGTTCTGCTGCTGACCTTGACCTAGTGCCTTAAAGCTATGCT 2873
QY 289 TATCTCAGCATGCTGCGCTTGGCAATGCTTATGTTTATGAGCGTATATATGAT 348
Db 2874 AGAGCGAGTATATATGAAACATTTATGTAAGCTTTTATATAGCTTTTACAGGTT 2933
QY 349 TGGAGCGCTATGCTTCGGAGAGCTGTTCTTGAATCTTTGCTTATCTGGCAGCTCT 408
Db 2934 GCCCGGTTTATGCGCATTTGGCTCTTCATATATATCTCATTAACCTTAAATATA 2993
QY 409 ACAATATTTGATGCGCACTCACCTTTCAGAGCTGCTGGATGTTCTTGGTATGG 468
Db 2994 CGCTATCTCAATGTAACATTTGACCTTCCAGGATTTGCGATTTCTTCAATAGG 3053
QY 469 GATGCGGTATGCAAAATGTTCTGTATTCGAAGAATCCGAGGAATTTTATATGTC 528
Db 3054 TATGCTGTGATACCAACATCTTAATTTTGAAGGTTCAAGAGAGCTGTGGGACGG 3113
QY 529 TCAAGTCTTAAAAATCTGTAGAAAAAGATATACCAAGCTTTTGGACCAATTTTGA 588
Db 3114 GAACACATAGAGCGCTGACATGATGAGAGATTTGCAAGAGCTATGAGACAGTAATGA 3173
QY 589 TTCTAATTGACTACGATATTTGCTTACGACTCTTTCTTCTTCTTCTTCTTCTTCTT 648
Db 3174 TGCAGATGTCACCACTTAATTCGCGGATCATCTCTTTTACTTGGGTTCAAGAAACGT 3233
QY 649 TAAAGGCTTTGCTTGTGACATTTAGGAATTTTCTCTTCAATGTTTACGGTCTTT 708
Db 3234 AAAAGTTTGTCTGTGACTCTTTAAAGAGTGTCTTCAAGCTTATTTACTGCTAATAC 3293
QY 709 CATGACATAATT-TTTCTCATGCTGTGATGATAAGACCAACATACAGATTTGCATA 767
Db 3294 TGATCAAGATTTCTTATGACAGCGCTCTTGAATGACCTTACAGAAATGCAATTT 3353
QY 768 TGATGAA 774
Db 3354 CTATGGA 3360
```

RBS

gene

RESULT 45
AE003876
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

AE003876
Xylella fastidiosa 9a5c, section 22 of 229 of the complete genome.
AE003876.1 GI:9105031

Xylella fastidiosa 9a5c.
Xylella fastidiosa 9a5c
Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
Xylella.

REFERENCE
1 (bases 1 to 11217)
Simpson, A.J., Reinach, F.C., Arruda, P., Abreu, F.A., Acencio, M.,
Alvarenga, R., Alves, L.M., Araya, J.E., Baia, G.S., Baptista, C.S.,
Barros, M.H., Bonaccorsi, E.D., Bordin, S., Bove, J.M., Briones, M.R.,
Bueno, M.R., Camargo, A.A., Camargo, L.E., Carraro, D.M., Carter, H.,
Colauto, N.B., Colombo, C., Costa, F.F., Costa, M.C., Costa-Neto, C.M.,
Coutinho, L.L., Cristofani, M., Dias-Neto, E., Docena, C., El-Dorri, H.,
Faccin, A.P., Ferreira, A.J., Ferreira, V.C., Ferro, J.A.,
Fraga, J.S., Franca, S.C., Franco, M.C., Frohme, M., Furlan, L.R.,
Garnier, M., Goldman, G.H., Goldman, M.H., Gomes, S.L., Gruber, A.,
Ho, P.L., Hoheisel, J.D., Junqueira, M.L., Kemper, E.L., Kitajima, J.P.
and Marino, C.L.

TITLE
The genome sequence of the plant pathogen Xylella fastidiosa. The
Xylella fastidiosa Consortium of the Organization for Nucleotide
Sequencing and Analysis

JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS

Nature 406 (6792), 151-157 (2000)
20365717
10910347

2 (bases 1 to 11217)
Simpson, A.J.G., Reinach, F.C., Arruda, P., Abreu, F.A., Acencio, M.,
Alvarenga, R., Alves, L.M.C., Araya, J.E., Baia, G.S., Baptista, C.S.,
Barros, M.H., Bonaccorsi, E.D., Bordin, S., Bove, J.M., Briones, M.R.S.,
Bueno, M.R.P., Camargo, A.A., Camargo, L.E.A., Carraro, D.M.,
Carter, H., Colauto, N.B., Colombo, C., Costa, F.F., Costa, M.C.R.,
Costa-Neto, C.M., Coutinho, L.L., Cristofani, M., Dias-Neto, E.,
Docena, C., El-Dorri, H., Faccin, A.P., Ferreira, A.J.S.,
Ferreira, V.C.A., Ferro, J.A., Fraga, J.S., Franca, S.C., Franco, M.C.,
Frohme, M., Furlan, L.R., Garnier, M., Goldman, G.H., Goldman, M.H.S.,
Gomes, S.L., Gruber, A., Ho, P.L., Hoheisel, J.D., Junqueira, M.L.,
Kemper, E.L., Kitajima, J.P., Krieger, J.E., Kuramae, E., Laigret, F.,
Lambais, M.R., Leite, L.C.C., Lenos, E.G.M., Lenos, M.V.F., Lopes, S.A.,
Lopes, C.R., Machado, J.A., Machado, M.A., Madeira, A.M.B.N.,
Madeira, H.M.F., Marino, C.L., Marques, M.V., Martins, E.A.L.,
Martins, E.M.F., Matsukuma, A.Y., Menck, C.F.M., Miracca, E.C.,
Miyaki, C.Y., Monteiro-Vitorello, C.B., Moon, D.H., Nagai, M.A.,
Nascimento, A.L.T.O., Netto, L.E.S., Nhani Jr., A., Nobrega, F.G.,
Nunes, L.R., Oliveira, M.A., de Oliveira, M.C., de Oliveira, R.C.,
Palmieri, D.A., Paris, A., Peixoto, B.R., Pereira, G.A.G., Pereira
Jr., H.A., Pesquero, J.B., Quaggio, R.B., Roberto, P.G., Rodrigues, V.,
de M. Rosa, A.J., de Rosa Jr., V.E., de Sa, R.G., Santelli, R.V.,
Sawasaki, H.E., da Silva, A.C.R., da Silva, P.R., da Silva, A.M., Silva
Jr., W.A., da Silveira, J.P., Silvestri, M.L.Z., Siqueira, W.J., de
Souza, A.A., de Souza, A.P., Terenzi, M.F., Truffi, D., Tsai, S.M.,
Teshako, M.H., Vallada, H., Van Sluys, M.A., Verjovski-Almeida, S.,
Vettore, A.L., Zago, M.A., Zatz, M., Meidanis, J. and Setubal, J.C.

Direct Submission
Submitted (02-JUN-2000) Organization for Nucleotide Sequencing and
Analysis, Bioinformatics Lab - IC/Unicamp, C.P. 6176, Campinas, SP
13083-970, Brazil

FEATURES
source
Location/Qualifiers
1. .11217
/organism="Xylella fastidiosa 9a5c"
/db_xref="taxon:160492"
/clone="9a5c"
229. .1440
/gene="9a5c"
229. .1440
/gene="XF0223"
229. .1440
/gene="XF0223"
/note="similar to SP|Q54177 (percent identity: 61 %/query
alignment coverage: 92.6 %/subject alignment coverage:
99.5 %); identified by sequence similarity; putative; ORF
located using Glimmer/RBSfinder"

gene
CDS

codon start=1
/transl_table=11
/product="queuine tRNA-ribosyltransferase"
/protein_id="AAF83036.1"
/db_xref="GI:9105032"
/translation="MGVLFVIRNDNRPLSCACRMSRLQFQLOQATDGAARRGQISFPFG
TVQPTAPMPVGTGAVKGLQCDLGAIEIILGNTFHLFRLPGLVEIADHGGHLSGDI-
RWNPILTDGGGQVFSLAHRRKISQGVTFAPATDGAQVFLGPESMKIQVLSGDI-
WIFDECTYPATEDVAGDMSLSLWRAQSRDAHDALDNDALFGLIOGVVPLDLRG
RSLDGLQALGFDGQYGGIAGVSESESRNVILEYHLPRVPADPRPRLMGVGRDVLVE
SVARGVDMFDCVMPTRHARNGQYFTGVTGKIRNACVARDVDIEFGCGGCPACVGGYT
RAYLRHLDRCNEMLASMLGLHNLWYETLMANMRAAITAGTFFAFRRSFYLRGLDL
PFLPEVAGCAG"
1551. .1913
/gene="XF0224"
1551. .1913
/genes="XF0224"
/note="similar to SP|P19677 (percent identity: 36 %/query
alignment coverage: 90.0 %/subject alignment coverage:
98.2 %); identified by sequence similarity; putative; ORF
located using Glimmer/RBSfinder"
codon start=1
/transl_table=11
/product="preprotein translocase YajC subunit"
/protein_id="AAF83037.1"
/db_xref="GI:9105033"
/translation="MEITMNLDFRIPVAQAQSAAPAGASPGGGLQMFPLPVLIIAM
YFSIIRPQMKRQKEYKSMMDHLGCGDEVITAGIAGVSVIGDEFVTVIAENVRVY
QKSAKALPKGSLPSAK"
1980. .3824
/gene="XF0225"
1980. .3824
/genes="XF0225"
/note="similar to SP|P19673 (percent identity: 44 %/query
alignment coverage: 100.0 %/subject alignment coverage:
99.8 %); identified by sequence similarity; putative; ORF
located using Glimmer/RBSfinder"
codon start=1
/transl_table=11
/product="protein-export membrane protein"
/protein_id="AAF83038.1"
/db_xref="GI:9105034"
/translation="MLEIPRMKYVLIFFVFLVLTLYALPNVYOKDP-QITANRGGH
DDAFQVMTDQAGSIVPKAVKEGESIMVRLMSLEMTQTRADYIAKVGENTVAL
NLASVPLWKLKGRPMVLGLDLGVGHAFALQVQKALNCKMFEVEVRVTLGSK
RIGFMRVRRPNTIRVTFDDAHARWMLFKQVGLNYSVDTAEAKRLIGATAT
QIASGAVSONLTTLRINELGVAEPIIQRQGDRIIVVELPGVSGDEWVNAVTVGDE
LEFRGVGDGDAYEAVRTGNVPEAKVYIRGTNPVLLNKR-VSGDEWVNAVTVGDE
LSPTRIAGVFGKNTTGLEAKENSLKLPAGSLAAP-FVEEYKVGRLKEEA
NGMPAVNTLSNVAQGMIDYTORNLKLMVYVIERPT-MVDEKVRGRLKEEA
RGVTVAVSEFLPTLVFTYVYRMFGVITSLALLFNLLAVMSLFGATMTLPGFAGL
ALVGLSDVANVLINERIEELRHGMPPRSALVAGYDTRKULKSLAI"
3844. .4809
/gene="XF0226"
3844. .4809
/genes="XF0226"
/note="similar to GI|1773093 (percent identity: 38 %/query
alignment coverage: 87.5 %/subject alignment coverage:
83.1 %); identified by sequence similarity; putative; ORF
located using Glimmer/RBSfinder"
codon start=1
/transl_table=11
/product="protein-export membrane protein"
/protein_id="AAF83039"
/db_xref="GI:9105035"
/translation="WKLFPADQVQRVLEGLGNAQVSGGNNELIRLQPYABH
NVALEFTGTGLVTRTQKFPAGNPATVQPGFVQVQKLANGVYATFIMLGFLTY
NNDVAKNAEAVRKV/DLILTAYMSLDEEDLTVLGALLSVMGFAINDIIVF
IGFRFENKFAVAASLNRINSQTLRSRTVAVMPFLSAFALYLVGGRSMEGLAETH
DRVRENFSRLSRVPLTIGLVKSKQDLLPKAKDLESLERRP"
5058
complement(

```

CDS
/ gene="tRNA-Glu-1"
/ complement(4984..5058)
/ gene="tRNA-Glu-1"
/ product="tRNA-Glu"
/ note="found by tRNAscan"
5139..6404
/ gene="XF0227"
5139..6404
/ gene="XF0227"
/ note="similar to SP|P31685 (percent identity: 48 %/query
alignment coverage: 97.9 %/subject alignment coverage:
87.3 %); identified by sequence similarity; putative; ORF
located using Glimmer/RBSfinder"
/ codon_start=1
/ transI_table=11
/ product="polynucleotide adenylyltransferase"
/ protein_id="AAB83040.1"
/ db_xref="GI:9105036"
/ translation="MLCVPRGHHTTSRKDISPNALVLYRLREAGFGALVYGAVRD
LIVGHPKDFDLATDAPDQIKQIFRSCLIGRFLAHVFGREIIIVATPERANSDD
GSGDRELENGRLVCDNVYGTIEDATRPDTCNALYTTEDFSDVDYGGFEDVQSL
MRIGDPRVRYOEDPVMLRAVLAVALGFEIAGAEPISSMAGLADVAAPALFEE
VLRPLSGHARVAPEALDEPGLIDVLPPSAKALNRSGALRVLIEGHSTDKKVI
KEBPVSPALFALLMPGFCHALVTTLQGVQVEBQRPAADPTVHLHRLVLPFRF
SLPMOEWLQNRVLSRQPKRVRLBHEFRFAFPLVRLAASHADVEFMKVA
OQOPSEDSLASAIDAVRFESEVSVSRQ"
6404..6922
/ gene="XF0228"
6404..6922
/ gene="XF0228"
/ note="similar to SP|P43777 (percent identity: 50 %/query
alignment coverage: 86.6 %/subject alignment coverage:
93.1 %); identified by sequence similarity; putative; ORF
located using Glimmer/RBSfinder"
/ codon_start=1
/ transI_table=11
/ product="2-amino-4-hydroxy-6-
hydroxymethylidihydropteridine pyrophosphokinase"
/ protein_id="AAB83041.1"
/ db_xref="GI:9105037"
/ translation="MPRSSRVVSGNVAVVAGLANLGPAMTLRAIALDOLPLTR
LRTASSLIRTPPMGRQDAEFVNAVALLETGLAPLALDLAEVCFGRVRLPGDRW
GPRTLDDLALYEQVLDLPRLRVPHPLHVRAPVVPVLAIEALVLPKGMVCELL
EHIDTLGLV2IR"
7033..7773
/ gene="XF0229"
7033..7773
/ gene="XF0229"
/ note="similar to SP|P31057 (percent identity: 55 %/query
alignment coverage: 98.8 %/subject alignment coverage:
100 %); identified by sequence similarity; putative; ORF
located using Glimmer/RBSfinder"
/ codon_start=1
/ transI_table=11
/ protein_id="3-methyl-2-oxobutanoate
adenylyltransferase"
Query Match 5 Score 98.2; DB 1; Length 11217;
Best Local Similarity 49.6 Pred. No. 1.5e-11;
Matches 262; Conservative 7; Mismatches 273; Indels 0; Gaps 0;

```

```

QY 352 AGCCGATCGCTGGAGAGCTGTTCTTGTAACTTTTGCTATCTGGAGAGCTCTACA 411
DB 3401 TGGTGTGATTAATCTCAATGGCACTGCTGTTTAATTTGTTGATGCGGTATGTC 3460
QY 412 GTATTGATGCGCACATCACCCTGTGAGACTCGCTGGATTTGTTGCTATGGGAT 471
DB 3461 GCTGTTGGTGCACACATGACCTTGCCGGGTTTGCCGGGTTGGCGCTGCGGT 3520
QY 472 GGGCGTAGATGCAATGTTCTTGATTTGAAAGAAATCCGAGAGAAATTTATGTC 531
DB 3521 TTGGGTGATGCTTAACGTGTTGATCAACGAGCGTATCCGTGAGAGTTGCGTAC 3580
QY 532 AGCTTAATAAATCTGTAGAAAAGATATPACAAAGCTTTTGAGACCATTTGATTC 591
DB 3581 GCCTCGAGGTGCGGATGTTGCGCGGTTCAGACAAAGCGGAGACGATTTGATGC 3640
QY 592 TAACCTGACTACAGTATTTGGCTCAGACACTTCTTCTTCTTCTTCTTCTTCTT 651
DB 3641 CAATTTGACCGGACGATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGA 3700
QY 652 AGGTTTGGCTTTGACATTTATTTAGAAATTTTCTTCAATGTTTACGGCTCTT 706
DB 3701 GGGCTTTGGCTTGAATGATGATGATGATGATGATGATGATGATGATGATGATG 3755

```

Search completed: December 17, 2002, 09:09:41
Job time : 9764 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model
Run on: December 17, 2002, 11:36:16 ; Search time 46.5 Seconds
(without alignments)
10682.974 Million cell updates/sec

Title: US-09-868-987-1
Perfect score: 3312
Sequence: 1 atggacttcgcataatgtc.....atgcacaattcagtaataatgc 1864

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 1816940

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ n2p.model -DEV=xlh
-Q=/cgn2.1/USPTO.spool-US09868987/runat.16122002.083005.23518/app_query.fasta_1.2055
-DB=A.GeneSeq.101002 -QFMT=fastan -SUFFIX=rag -MINMATCH=0 -1-LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=45
-MODE=LOCAL -OUTFMT=pco -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09868987.qcgn.1.1.4 @runat.16122002.083005.23518 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -LONGLOG -DEV TIMEOUT=120
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FCGAPOP=6 -FCGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : A.GeneSeq.101002.*

- 1: /SID22/gcgdata/geneSeq/geneSeq-embl/AA1980.DAT.*
- 2: /SID22/gcgdata/geneSeq/geneSeq-embl/AA1981.DAT.*
- 3: /SID22/gcgdata/geneSeq/geneSeq-embl/AA1982.DAT.*
- 4: /SID22/gcgdata/geneSeq/geneSeq-embl/AA1983.DAT.*
- 5: /SID22/gcgdata/geneSeq/geneSeq-embl/AA1984.DAT.*
- 6: /SID22/gcgdata/geneSeq/geneSeq-embl/AA1985.DAT.*
- 7: /SID22/gcgdata/geneSeq/geneSeq-embl/AA1986.DAT.*
- 8: /SID22/gcgdata/geneSeq/geneSeq-embl/AA1987.DAT.*
- 9: /SID22/gcgdata/geneSeq/geneSeq-embl/AA1988.DAT.*
- 10: /SID22/gcgdata/geneSeq/geneSeq-embl/AA1989.DAT.*
- 11: /SID22/gcgdata/geneSeq/geneSeq-embl/AA1990.DAT.*
- 12: /SID22/gcgdata/geneSeq/geneSeq-embl/AA1991.DAT.*
- 13: /SID22/gcgdata/geneSeq/geneSeq-embl/AA1992.DAT.*
- 14: /SID22/gcgdata/geneSeq/geneSeq-embl/AA1993.DAT.*
- 15: /SID22/gcgdata/geneSeq/geneSeq-embl/AA1994.DAT.*
- 16: /SID22/gcgdata/geneSeq/geneSeq-embl/AA1995.DAT.*
- 17: /SID22/gcgdata/geneSeq/geneSeq-embl/AA1996.DAT.*
- 18: /SID22/gcgdata/geneSeq/geneSeq-embl/AA1997.DAT.*
- 19: /SID22/gcgdata/geneSeq/geneSeq-embl/AA1998.DAT.*
- 20: /SID22/gcgdata/geneSeq/geneSeq-embl/AA1999.DAT.*
- 21: /SID22/gcgdata/geneSeq/geneSeq-embl/AA2000.DAT.*
- 22: /SID22/gcgdata/geneSeq/geneSeq-embl/AA2001.DAT.*
- 23: /SID22/gcgdata/geneSeq/geneSeq-embl/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2868	86.6	997	20	AAV35184
2	2778	83.9	552	21	AAV35543
3	1593	48.1	1250	20	AAV37780
4	702	21.2	981	20	AAV34476
5	702	21.2	981	20	AAV34352
6	678.5	20.5	737	20	AAV30754
7	647.5	19.6	741	22	AAV32042
8	647.5	19.6	766	23	ABP38844
9	618.5	18.7	754	23	AB448641
10	496	15.0	126	20	AAV37781
11	431	13.0	525	19	AAV98248
12	426	12.9	526	18	AAV55688
13	408.5	12.3	616	22	AAV88551
14	408.5	12.3	616	23	AAU91482
15	364	11.0	618	23	AAU91065
16	359.5	10.9	773	22	ABG28914
17	357.5	10.8	713	22	ABG26197
18	351	10.6	323	19	AAV98768
19	349	10.9	140	20	AAV37782
20	342	10.3	325	22	AAV88556
21	342	10.3	325	23	AAU91487
22	306.5	9.3	576	22	AAU63631
23	297	9.0	311	23	AAU91065
24	277.5	8.4	614	22	AAV78899
25	277.5	8.4	637	23	AAE15862
26	277.5	8.4	644	22	AAV91576
27	245.5	7.4	279	21	AAV05947
28	234	7.1	418	22	AAU63575
29	228	6.9	403	22	AAV91574
30	228	6.9	403	23	AAE15863
31	147.5	4.5	1055	22	AAV36596
32	147	4.4	746	22	AAV96403
33	144	4.4	564	20	AAV35183
34	144	4.3	1034	22	AAU34738
35	140	4.2	178	22	AAV82041
36	140	4.2	744	23	ABV49954
37	139.5	4.2	507	22	AAV96433
38	139	4.2	1098	22	ABG30367
39	136.5	4.1	1007	20	AAV19957
40	136.5	4.1	1032	22	AAU35552
41	136.5	4.1	1036	20	AAV19956
42	136.5	4.1	1051	23	ABV40160
43	130.5	3.9	730	22	AAV90442
44	130.5	3.9	730	22	AAV78979
45	129	3.9	1040	21	AAV15925

ALIGNMENTS

RESULT 1
AAV35184
ID AAV35184 standard; Protein; 997 AA.
XX
AC AAV35184;
XX
DT 13-SEP-1999 (first entry)
XX
DE Chlamydia pneumoniae transport polypeptide.

Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis; sinusitis; purulent otitis media; erythema nodosum; pharyngitis;
vaccine; neutralising epitope.

Chlamydia pneumoniae.

WO9927105-A2.

03-JUN-1999.

```

XX PF 20-NOV-1998; 98MO-1B01830.
XX PR 04-NOV-1998; 98US-0107078.
XX PR 21-NOV-1997; 97FR-0014673.
XX (GSET ) GENSET.
PA
PI Griflais R;
XX
XX MPI; 1999-357842/30.
XX
XX Genome sequence of Chlamydia pneumoniae
XX
XX Page 1042-1044; Disclosure; 1912pp; English.
XX
XX AAY34584-Y35879 represent the proteins encoded by all the open reading
XX frames in the complete genome (see AAY91990) of Chlamydia pneumoniae.
XX C. pneumoniae causes respiratory disease such as pneumonia and
XX C. pneumoniae causes respiratory disease such as pneumonia and
XX bronchitis and is thought to be a contributing factor in heart
XX disease, sarcoidosis, sinusitis, purulent otitis media, erythema
XX nodosum or pharyngitis. The polypeptides encoded by the open reading
XX frames of the C. pneumoniae genome (see AAY34584-Y35879) can be used in
XX immunogenic compositions as vaccines. Vectors containing C. pneumoniae
XX nucleotide sequences can also be used as immunogenic compositions,
XX especially where the vector directs the expression of a neutralising
XX epitope of C. pneumoniae.
XX
XX SQ Sequence 997 AA;
XX
XX Alignment Scores:
XX Pred. No: 1.3e-315 Length: 997
XX Score: 2868.00 Matches: 583
XX Percent Similarity: 99.49% Conservative: 0
XX Best Local Similarity: 99.49% Mismatches: 3
XX Query Match: 86.59% Indels: 3
XX DB: 20 Gaps: 0
XX
XX US-09-868-987-1 (1-1864) x AAY35184 (1-997)
XX
XX 2 TGGACATTCGCGATATTTGTCAGAGGAGATCAGCGGACCTGCTAATGACAAATATTTCTGCA 61
XX DB 412 TTTPhSerAlaTyrGySGInGluGlyIleSerGlyThrAlaAsnGlyGlnTyrSerAla 431
XX
XX 62 AACGTCGATGCGCTATGCTGTAGTATGACGCTTATATGTCACAGCCCTATTTTA 121
XX DB 432 AsnArgIlyTrpArgMetAlaValAlaIleAspGlyTyrMetValSerSerProIleLeu 451
XX
XX 122 AACGTCGATGCGCTATGCTGTAGTATGACGCTTATATGTCACAGCCCTATTTTA 121
XX DB 452 AsnArgIlyTrpArgMetAlaValAlaIleAspGlyTyrMetValSerSerProIleLeu 471
XX
XX 182 AAATCGCCTCAGATTTAAATCTGAGCGATGCTTTTGTCCCGAGGTTCTCAGTGA 241
XX DB 472 LysLeuAlaSerIlePheLeuLysSerGlyAlaMetSerPheValProGlyValLeuSerGlu 491
XX
XX 242 GAGACGATCTCTTCTGATCTTTGGGAAAAACAATGTACACAGGCAATTATCTCAGATGC 301
XX DB 492 GluThrIleSerSerAspLeuGlyLysGlnCysThrGlnGlyIleIleSerAlaCys 511
XX
XX 302 TGTGGCTTTGGCAATGCTTATTTGATGAGCGTATTTATTAATTTGGAGGGCTCATC 361
XX DB 512 CysGlyLeuAlaMetIleValLeuMetSerValTyrTyrAspPheGlyGlyValIle 531
XX
XX 362 GCTTCGGGAGCTGTTCTTCTGAATCTTTGCTTATCTGGGAGGCTTACAGATTTTGGAT 421
XX DB 532 AlaSerIlyAlaValIleLeuAsnLeuLeuIleTrpAlaAlaLeuGlnTyrLeuAsp 551
XX
XX 422 GCGGCACTCACTTGTGAGACTGCTGGAGATTTGTTCTGTATGGGAGTGGCGGTGAT 481
XX DB 552 AlaProLeuThrIleuSerGlyLeuAlaGlyIleValLeuAlaMetGlyMetAlaValAsp 571
XX
XX 482 GCAATGTTCTTGTATTCGAAAGAAATCCGAGAGGAATTTTATGTCTCAAAAGTCTTAA 541

```

```

DB 572 AlaAsnValLeuValPheGluArgIleArgGlnGluPheLeuLeuSerGlnSerLeuLys 591
XX
XX 542 AAATCTGTAGAAAAAGATATACCAAGGCTTTTGAGCCATTTTGTATCTTAACTTGACT 601
XX DB 592 LysSerValGlyLysGlyTyrThrLysAlaPheGlyAlaIlePheAspSerAsnLeuThr 611
XX
XX 602 ACAGTATTTGGCCCTCAGACACTTCTTTCTCTAGATACAGGGCGCTTAAAGGTTTGCT 651
XX DB 612 ThrValLeuAlaSerAlaLeuLeuPhePheLeuAspThrGlyProIleLysGlyPheAla 631
XX
XX 662 TTGACATTTGATTTTAGAAATTTCTCTCAATGTTTACGCGCTTTTCTTACGACTAAATTT 721
XX DB 632 LeuThrLeuIleLeuGlyIlePheSerSerMetPheThrAlaLeuPheMetThrLysPhe 651
XX
XX 722 TTCTTCATGCTGTGATGATGATAAGACCAACATACACAGTTCATATGATAATGATTC 781
XX DB 652 PhePheMetLeuThrPheLeuAsnLysThrGlnHisThrGlnLeuHisMetMetLeuLysPhe 671
XX
XX 782 GTGGGATTAAGATGATTTCTTTGAGAGGATGCAAAAACTTTGGGCTGTTTCTGGAAGT 841
XX DB 672 ValGlyIleLysHisAspPheLeuArgGlyCysLysLeuThrPalAvalSerGlySer 691
XX
XX 842 GTTTTCTTTTAGGTTGCGGCTCGGGTTTGGAGCCGGAATCCGTTTGGGGAATG 901
XX DB 692 ValPheLeuLeuGlyCysValAlaLeuGlyPheGlyAlaTrpAsnSerValLeuGlyMet 711
XX
XX 902 GATTTTAAAGAGGATGATGCTTTTACCTTTAATCAAAAAGAGATGATCAGCGATGTT 961
XX DB 712 AspPheLysGlyGlyIlyTrAlaPheThrPheAsnProLysGlyLysIleSerAspVal 731
XX
XX 962 GCTCAATGCTGTGGCAAGTGTGCTAAACTACAGAGAGCTGCTTCTTCTTAAGAC 1021
XX DB 732 AlaGlnMetCysGlyLysValAlaHisLysLeuGlnGlnAlaGlyLeuSerSerArgAsp 751
XX
XX 1022 TTCGCTATTCAAACTTTGGATCTTCAGAAAAAGATCAAAATCATTTTATGTAATTAAGCT 1081
XX DB 752 PheArgGlyIleGlnThrPheGlySerSerGlyLysIleLysIleTyrPheSerAspLysAla 771
XX
XX 1082 TTAAGCTATATCAAGCA-GATACGAGCTCTCTCTTAAATTAACGATCATGAGCTGGCG 1140
XX DB 772 LeuSerTyrThrLysAlaSerThrSerLeuSerProLysIleAsnAspHisGlyLeuAla 791
XX
XX 1141 TTAAT- GTGGATTTGTTGTCAGAAACAAGGCTTATTTCTTACGGAACCTTAAACGA 1199
XX DB 792 LeuAlaValGlyLeuLeuSerGlyLysThrGlyLeuAspPheSerThrIleThrLeuAsnGly 811
XX
XX 1200 ACCGCAAAA-TTTTGTCAAAAGTAAAGCAGCAAACTATCGAAGAAATGCGTTATCAGCGC 1258
XX DB 812 ThrGlnAsnPheThrPheLysValSerSerLysLeuSerLysHisMetArgTyrGlnAla 831
XX
XX 1259 ACCATGGGCTTTTAAAGCTTTGGCAATCATCTTCTATGTAGTTTGGCTTTGA 1318
XX DB 832 ThrIleGlyLeuLeuGlyAlaLeuAlaIleIleLeuLeuTyrValSerLeuArgPheGlu 851
XX
XX 1319 TGGCAATATGCTTTCAAGTCCGATAGCGCTTTAATTCATGACCTTTTGGCTTACTGTGCA 1378
XX DB 852 TrpGlnTyrAlaPheSerAlaValCysAlaLeuIleHisAspLeuLeuAlaThrCysAla 871
XX
XX 1379 GTCTTTTATAGACATTTCTTTTGAAGAAATCAATAGATTTTGAAGGCAATTTGGT 1438
XX DB 872 ValLeuPheIleAlaHisPhePheLeuLysLysIleGlnIleAspLeuGlnAlaIleGly 891
XX
XX 1439 GCTTTAATGACTGATTTGGGATTCATTAACAAATACCTTGTATGATTTTGTATGCTATT 1498
XX DB 892 AlaLeuMetThrValLeuGlyTyrSerLeuAsnAsnThrLeuIleIlePheAspArgIle 911
XX
XX 1499 CGTGAAGATGGCAAGCAACCTGTTTACCCCTATGCAAGTTTTAACTTAATGATGCCCTT 1558
XX DB 912 ArgGluAspArgGlnAlaAsnLeuPheThrProMetHisValLeuValAsnAspAlaLeu 931
XX
XX 1559 CAAAAGCGTTGACCGGACCGGTAATGACAACAGCTACCACTTACAGTTTGTAAATG 1618

```

Db 932 GlnLysThrPheSerArgThrValMetThrThrAlaThrThrLeuSerValLeuLeuMet 951
Qy 1619 CTTTGTGTTATAGCGGCTCTCTGTTTAAATTTGCAATTTATGACCATAGGGATT 1678
Db 952 LeuLeuPheIleGlySerSerValPheAsnPheAlaPheIleMetThrIleGlyIle 971
Qy 1679 CTTCTAGAACTTTATCTCTCTTTATATATGACCACTCTGTTGTTGTTATGTCGGT 1738
Db 972 LeuLeuGlyThrLeuSerSerLeuTyriLeAlaProProLeuLeuPheMetValArg 991
Qy 1739 AAAGAAATCGTCAAAA 1756
Db 992 LysGluAsnArgSerLys 997

RESULT 2
AA95543
ID AA95543 standard; Protein; 552 AA.
XX AC
XX AC AA95543;
XX 10-OCT-2000 (first entry)
XX Chlamydia pneumoniae antigen CPN100686 RY 54.
XX CPN100686 RY 54; antigen; infection; diagnosis; therapy; vaccine;
XX outer membrane protein.
XX Chlamydia pneumoniae.
XX OS
XX WO200039158-A1.
XX 06-JUL-2000.
XX 23-DEC-1999; 99WO-CA01230.
XX 23-DEC-1998; 98US-0113280.
XX 23-DEC-1998; 98US-0113281.
XX 23-DEC-1998; 98US-0113282.
XX 23-DEC-1998; 98US-0113283.
XX 23-DEC-1998; 98US-0113284.
XX 23-DEC-1998; 98US-0113285.
XX 23-DEC-1998; 98US-0113385.
XX 28-DEC-1998; 98US-0114050.
XX 28-DEC-1998; 98US-0114056.
XX 28-DEC-1998; 98US-0114057.
XX 28-DEC-1998; 98US-0114058.
XX 28-DEC-1998; 98US-0114059.
XX 28-DEC-1998; 98US-0114061.
XX (CONN-) CONNAUGHT LAB LTD.
XX Murdin AD, Oomen RP, Wang J;
XX WPI; 2000-452369/39.
XX N-PSDB; AAA50030.
XX Novel Chlamydia polynucleotides and polypeptides useful for diagnosis,
XX prevention and treatment of Chlamydia infection in mammals -
XX Claim 16(a); Page 55-58; 215pp; English.
XX The present sequence is that of Chlamydia pneumoniae antigenic
XX protein CPN100686 RY 54, a putative 93 kDa outer membrane protein.
XX It is an example of C. pneumoniae antigenic polypeptides of the
XX invention (see AA95543-55) that are encoded by open reading frames
XX (see AAA50030-42) identified in the C. pneumoniae genome. The
XX polypeptides are useful in the diagnosis, treatment and prevention
XX of Chlamydia infection. They can be prepared by recombinant
XX methods using transformed unicellular host cells. The
XX polypeptides, or immunogenic fragments of them, or fusion proteins
XX that include an additional polypeptide such as a heterologous
XX signal peptide or a polypeptide having adjuvant activity, are
XX utilised as vaccines to protect against Chlamydia infection.

CC The polypeptides are also used in diagnostic kits and in methods
CC of detecting Chlamydia infection.
XX
SQ Sequence 552 AA;
Alignment Scores:
Pred. No.: 1.53e-305 Length: 552
Score: 2778.00 Matches: 552
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 83.88% Indels: 0
DB: 21 Gaps: 0
US-09-868-987-1 (1-1864) x AA95543 (1-552)
Qy 101 ATGGTCAGCAGCCCTATTTTAAACGTCCTCCATTTGAAATATCATGCAGTGTCTCAGGGAAA 160
Db 1 MetValSerSerProIleLeuAsnValProLeuLysAsnHisAlaSerValSerGlyLys 20
Qy 161 TTTACCCACCGTGAAGTGAGCAAACTCGCCTCAGATTTAAATCTGGAGCGATGCTTTT 220
Db 21 PheThrHisArgGluValSerLysLeuAlaSerAspLeuLysSerGlyAlaMetSerPhe 40
Qy 221 GTTCCCGAGGTTCTCAGTGAAGACAGCATCTTCTCATCTTGGGAAAAACAATGTACA 280
Db 41 ValProGluValLeuSerGluThrIleSerSerAspLeuGlyLysGlnCysThr 60
Qy 281 CAAGGCATTATCTCAGCATGCTGTGGCTTGGCAATGCTATTGTTTGTATGAGCGTATAT 340
Db 61 GlnGlyIleIleSerAlaCysCysGlyLeuAlaMetLeuIleValLeuMetSerValTy 80
Qy 341 TATAGATTGGAGCGCTCATCGCTTCGGGAGCTGTTCTTCTGAATCTTTTCTTATCTGG 400
Db 81 TyrArgPheGlyGlyValIleAlaSerGlyAlaValLeuLeuAsnLeuLeuIleTrp 100
Qy 401 GCAGCTCTACAGTATTTGGATGCGCCACTACCTTGTCTCAGGACTCGCTGGGATTTGTTCT 460
Db 101 AlaAlaLeuGlnTyrLeuAspAlaProLeuThrLeuSerGlyLeuAlaGlyIleValLeu 120
Qy 461 GCTATGGGATGGCGGTAGATGCAAAATGTTCTTCTTATTCGAAAGAATCCGAGAGGAATT 520
Db 121 AlaMetGlyMetAlaValAspAlaAsnValLeuValPheGluArgIleArgGluGluPhe 140
Qy 521 TTATTGTCTCAAGCTTTAAAAAATCTGTAGAAAAAGATATACCAAGGCTTTTGGAGCC 580
Db 141 LeuLeuSerGlnSerLeuLysSerValGluLysGlyTyrThrLysAlaPheGlyAla 160
Qy 581 ATTTTGAATCTAACTTGACTACAGTATTTGGCCTCAGCATTTCTTTCTTCTCTAGATACA 640
Db 161 IlePheAspSerAsnLeuThrThrValLeuAlaSerAlaLeuLeuPhePheLeuAspThr 180
Qy 641 GGGCTATTAAAGGTTTGTCTTGACATTTTGTAGGAATTTTCTCTCAATGTTTACG 700
Db 181 GlyProIleLysGlyPheAlaLeuThrLeuIleLeuGlyIlePheSerSerMetPheThr 200
Qy 701 GCTCTTTTTCATGACTAAATTTTCTTTCATGCTGTGGATGAATAGACCCCAACATACACAG 760
Db 201 AlaLeuPheMetThrLysPhePheMetLeuTrpMetAsnLysThrGlnHisGln 220
Qy 761 TTGCATATGATGAATAAGTTTCGTGGGATAAAGCATGATTTCTTGAGAGGATGCAAAAAA 820
Db 221 LeuHisMetMetAsnLysPheValGlyIleLysHisAspPheLeuArgGlyCysLysLys 240
Qy 821 CTTTGGGCTGTTTCTGCAAGTGTCTTTCTTTTGTAGTTGCGTGTCTCGGTTTGGAGCC 880
Db 241 LeuTrpAlaValSerGlySerValPheLeuLeuGlyCysValAlaLeuGlyPheGlyAla 260
Qy 881 TGGAAATTCGTTTGGGAATGGATTTTAAAGAGGGGTATGCTTTTACCTTTAAATCCAAAA 940
Db 261 TrpAsnSerValLeuGlyMetAspPheLysGlyGlyTyrAlaPheThrPheAsnProLys 280
Qy 941 GAGCATCGCATCAGCGATGTTGCTCAAAATCGTGGCAAAAGTTGTGCATAAATCTACAGAA 1000

Db 281 GluHisGlyIleSerAspValAlaGlnMetArgIlyIysValValHisIlySleuGlnIu 300
 Qy 1001 GCTGGCTTTCTTCTAGAGATTCCGTAATCAAACTTTGGATCTTCAGAAAGATCAAA 1060
 Db 301 AlAGlyLeuSerSerArgAspPheArgIleGlnThrPheGlySerSerGlyIlySle 320
 Qy 1061 ATCTATTTTGTGTAAAGCTTTAAGCTATATCTAAGCAAGATAGAGGCTCTCTCAAA 1120
 Db 321 IleTyrPheSerAspIlySleuSerTyrThrIySglnIleAlaGlaSerLeuLeuIys 340
 Qy 1121 TTAAAGATCATGAGCTGGCGTTATTGTGGATGTGTTCAGAAAACAGGCTAGATTCTC 1180
 Db 341 LeuThrIleMetSerTrpArgTyrCysGlyIleValValAlaGlnAsnArgProArgPheLeu 360
 Qy 1181 TACGGAACCTCTTAACGAAACGAAATTTTGTCAAGAGTACAGCAAACTATCGAAG 1240
 Db 361 TyGlyIysSerIlySAspAsnAlaIlyPheTrpSerIlySAspSerIlySLeuSerIys 380
 Qy 1241 AAAATGCGTTATACGGGACCATCGGCTTTTAGAGCTTTGGCAATCATCTGCTCAT 1300
 Db 381 LysMetArgTyrGlnAlaThrIleGlyLeuLeuGlyAlaLeuAlaIleIleLeuLeuTyr 400
 Qy 1301 GTGAGTTGGCGCTTGAATGCGCAATATGCTTTCAGTCCGATGCGCTTAAATCATGAC 1360
 Db 401 ValSerLeuArgPheGlnTyrPglntTyrAlaPheSerAlaValCysAlaLeuIleHisAsp 420
 Qy 1361 CTTTGGCTACCTGTGACGCTTGTGTATAGACATTTCTTTTGAAGAAATTCAAATA 1420
 Db 421 LeuLeuAlaThrCysAlaValLeuPheIleAlaHisPheLeuIlySleIleGlnIle 440
 Qy 1421 GATTGCAAGCGATTTGGCTTTTATAGACTGATTTGGGATTTGATTAACAAATCTTG 1480
 Db 441 AspLeuGlnAlaIleGlyAlaLeuMetThrValLeuGlyTyrSerLeuAsnThrLeu 460
 Qy 1481 ATCATTTTGTATCGATTCCTGAGAGATGCGCAAGCAACGTTTACCCCTATCATGTT 1540
 Db 461 IleIlePheAspArgIleAlaGlnPheArgGlnAlaAsnLeuPheThrProMetHisVal 480
 Qy 1541 TTACTTATGATGCGCTTCMAAAGACGTTACGCGCAGGATATGACAAACGATCAACT 1600
 Db 481 LeuValAsnAspAlaLeuGlnLysThrPheSerArgThrValMetThrThrAlaThrThr 500
 Qy 1601 CTATCAGTTTGTATATGCTTTTGTATAGCGCGCTCTCTGCTTAAATTTGCAATT 1660
 Db 501 LeuSerValLeuLeuMetLeuLeuPheIleGlyIlySerSerValPheAsnPheAlaPhe 520
 Qy 1661 ATTATGACCATAGGAGATCTCTAGGAACCTTATCGTCTTTTATATGACCACTCTG 1720
 Db 521 IleMetThrIleGlyIleLeuLeuGlyThrLeuSerSerIlyIleAlaProProLeu 540
 Qy 1721 TTGTTGTTATGTCCTGTAAGAAATGCTCAAA 1756
 Db 541 LeuLeuPheMetValArgIlySglnAsnArgSerIys 552
 RESULT 3
 AAY37780
 ID AAY37780 standard; Protein; 1250 AA.
 XX AAY37780;
 DT 07-OCT-1999 (first entry)
 DE Chlamydia trachomatis outer membrane protein.
 XX
 XX Vaccine; eye disease; conventional trachoma; nonendemic trachoma;
 KM paratrachoma; inclusion conjunctivitis; genital disease; perithenitis;
 KM nongonococcal urethritis; epididymitis; cervicitis; salpingitis;
 KM Bartholinitis; pneumonia; venereal lymphogranulomatosis.
 XX
 OS Chlamydia trachomatis.
 XX
 PN W09928475-A2.
 XX

PD 10-JUN-1999.
 XX
 PF 27-NOV-1998; 98WO-IB01939.
 XX
 PR 04-NOV-1998; 98US-0107077.
 PR 28-NOV-1997; 97PR-0015041.
 PR 17-DEC-1997; 97PR-0015034.
 XX
 PA (GENSET) GENSET.
 XX
 PI Griffats R;
 XX
 DR WPI; 1999-371125/31.
 XX
 PT Genome sequence of Chlamydia trachomatis
 PS Disclosure; Page 1365; 1755pp; English.
 XX
 CC AAY36754-Y37949 are encoded by open reading frames (ORFs) of the genome
 CC of Chlamydia trachomatis (see A201425). The polypeptides can be used as
 CC vaccines against Chlamydia trachomatis. Antisense and ribozyme sequences
 CC can also be used to control growth of the microorganism. Chlamydia
 CC trachomatis is responsible for a large number of diseases, e.g. eye
 CC diseases such as conventional trachoma, nonendemic trachoma,
 CC paratrachoma, and inclusion conjunctivitis; genital diseases such as
 CC nongonococcal urethritis, epididymitis, cervicitis, salpingitis,
 CC perithenitis, Bartholinitis; pneumonia; venereal lymphogranulomatosis;
 CC and venereal lymphogranulomatosis. The polypeptides of the invention
 CC may be of use in treating these diseases.
 XX
 SQ Sequence 1250 AA;
 Alignment Scores:
 Pred. No.: 6.17e-171 Length: 1250
 Score: 1593.00 Matches: 319
 Percent Similarity: 75.96% Conservative: 57
 Best Local Similarity: 64.44% Mismatches: 73
 Query Match: 48.10% Indels: 46
 DB: 20 Gaps: 5

US-09-868-987-1 (1-1864) x AAY37780 (1-1250)
 Qy 2 TGGACCTCCGCAATTTGTCAGAGGGGATCAGCGGCACTGTAATGCAATATTCGCA 61
 Db 776 TrpThrSerIysPheCysGlnGlnIlyValSerGlyThrIySAsnGlyLeuPheSerGly 795
 Qy 62 AACCGTGGATGGCGTATGCGTGTAGTATGACGTTATATGATGACAGCCCTATTTTA 121
 Db 796 GlyArgGlyTrpArgMetAlaValIleLeuAspGlyTyrValIleSerAspProValLeu 815
 Qy 122 AACGTCCCATGAAATATCGCAGTGTCTCAGGGAATTTACCCACCGTGAAGTACG 181
 Db 816 AsnValProLeuIysAspHisAlaSerValSerGly***PheSerTyrArgGlnValHis 835
 Qy 182 AAACCTGGCCCTCAGATTTAAATCTGGAGCATGCTTTTGTCCCGAGTCTCAGTGA 241
 Db 836 ArgLeuAlaThrAspLeuIysSerGlyAlaMetSerPheIleProGlnIleLeuSerGln 855
 Qy 242 GAGACGATCTCTTCTGTATCTTGGAAAAAACAATGTAACACAGCATATATCGACATGC 301
 Db 856 GluValValSerProGlnLeuGlySerSerGlnArgValGlnGlyIleLeuSerValVal 875
 Qy 302 TGTGGCTGGCAATGCTTATATGTTTGTATGATGAGCGTATATATATGATTTGAGCGCTATC 361
 Db 876 LeuGlyLeuValValIleLeuIleValIleMetSerValTyrTyrArgPheGlyIlyValIle 895
 Qy 362 GCTTCGGAGACGTGCTTCTGGAATCTTTGCTATATCGGAGCTCATACAGTATTTGGAT 421
 Db 896 AlaSerIleAlaValIleLeuAsnLeuLeuLeuIleTrpAlaSerMetGlnTyrLeuAsp 915
 Qy 422 GCGCCACTCACTTGTTCAGAGCTCGCTGGATGTTCTTCTATGCGGATGCGCGTATGAT 481
 Db 916 AlaProLeuThrLeuSerSerGlyLeuAlaGlyIleIleLeuAlaMetGlyMetAlaValAla 935

Db 459 GlyAspLeuAlaIasnValLeuAsnSerGlyLysMetSerAspAlaThrValSerIleGluIn 478
 QY 239 GAAGAGACGATCTTCTGTGATCTTGGGAAAAACATGTAACAAGCATTATCTCGCA 298
 Db 479 GluAsnValIleGlyProThrLeuGlyAlaGlySerIleLysAlaGlyPheLeuSerPhe 498
 QY 299 TGCCTGGCTTGGCAATGCTTATTTGTTATGAGCGCTATATTATTAATTT---GGAGGC 355
 Db 499 LeuLeuAlaLeuValIleLeuMetCysTyrMetCysLeuAlaTyrGlyPheLeuProGly 518
 QY 356 GTCATCGCTTCGGAGCGTGTCTCTGATCTTGTGATCTTGTGCGGACGCTACAGTAT 415
 Db 519 LeuIleAlaIasnGlyAlaLeuIleValAlaSerPhePheThrLeuGlyValIleAlaSer 538
 QY 416 TTGGATGGCCACTACCTTATTCAGACTCGCTGGATGTTGTTGCTTATGGGATGGCC 475
 Db 539 PheHisAlaValLeuThrLeuSerGlyIleAlaGlyLeuValLeuThrLeuGlyMetAla 558
 QY 476 GTGATGCAATGTTCTTGTGATTCGAAAGATCCGAGAGAAATTTTATTTGTCCTCAAGT 535
 Db 559 ValAspAlaIasnValLeuIlePheGluArgIleLysGluLeuAlaGlyLysThr 578
 QY 536 CTTAAAAATCTGTAGAAAAAGATATACCAAGCCTTGGAGCGCATTTTGTGATTTTAAAC 595
 Db 579 ProIleArgAlaValThrAspGlyTyrGlyAsnAlaPheSerAlaIlePheAspSerAsn 598
 QY 596 TTGACTACAGTATTTGGCTCGACACTTCTTCTTCTCTAGATACAGGCGCTATTAAAGG 655
 Db 599 ValThrThrIleIleThrGlyIleIleLeuPheLeuTyrGlyThrGlyProIleArgGly 618
 QY 656 TTTCCTTGCATTTGATTTAGAAATTTCTCTTCATGTTTACGGCTTTTCATGACT 715
 Db 619 PheAlaThrThrLeuIleIleGlyLeuIleAlaSerPheIleThrAlaValPheLeuThr 638
 QY 716 AAATTTTCTTCATGCGCTGTG-----ATGATTAAGCCCAATACACAG 760
 Db 639 ArgIleValPheGlyLysLeuAlaLysLysGlyArgLeuAspLysIleThrPheThrThr 658
 QY 761 TTGCATATGATGAATTAATTTGCTGGGATTAAGCATGATTTTTCAGAGATGCAAAAA 820
 Db 659 SerIleThrArgAsnLeuLeuValAsnProSerTyrAsnIleLeuGlyLysArgLysThr 678
 QY 821 CTTTGGCGTCTTTCGAAAGTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 871
 Db 679 GlyPheIleIleProValIleIleIleValLeuGlyLeuIleAlaSerPheThrIleGly 698
 QY 872 TTTGGAGCTGGAATTCGCTTTTGGGATGATTTTAAAGAGG-----TATGCTCTT 925
 Db 699 LeuAsnArg-----GlyIleGluPheSerGlyGlyArgAsnTyrValVal 713
 QY 926 ACCTTTAATCCAAAAGACATGCGATTCAGCATGTTGCTCAATGCTGGCAAGTTGTG 985
 Db 714 LysPhe-----AspGlnProValSerSerGluAla---ValArgSerAlaLeuSer 729
 QY 986 CATAAACTACGGA-----GCTGCTCTTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1027
 Db 730 SerProLeuGlnGluLysValLeuValThrSerIleGlyThrGluGlyThnGluValArg 749
 QY 1028 ATTCAACA-----TTTGGATCTTCAGAAAAGATCAAAATCTATTATTAGTAT 1075
 Db 750 IleSerThrAsnTyrLysIleGlnGluSerGluGluThrGluAlaGluIleThrAsp 769
 QY 1076 AAAGCTTTAAGC-----TATACTAAGCAGATACAGACCTCTCTCTTAAAA 1120
 Db 770 LysLeuTyrGlnSerLeuLysGlyPheTyrThrGlnGlnProThrAlaAspGlnPheLeu 789
 QY 1121 TTAACGATCATGAGCTGCGCTTATTTGTGGATTTGTTCAGAAACAGCCTTAGATTCTC 1180
 Db 790 AspAsnIleIleSerSerGln----- 796
 QY 1181 TAGCGAAACTTAAGCAAGCAAAATTTTGTCAAGGTAGACAGCAAACTATCGAAG 1240
 Db 797 -----LysValSerProSerMetSerSer 804

QY 1241 AAAATGCTTATCAGCGCACATCGGCTTTTATGAGACTTTGGCAATCATCTTGTCTAT 1300
 Db 805 AspIleThrArgGlyAlaIleThrAlaValLeuLeuSerMetIlePheMetAlaIleTyr 824
 QY 1301 GTGAGTTTGGCTTTGATGGAATGGAATGCTTCAAGTCC-----GTATGC 1345
 Db 825 IleLeuIleArgPheArg---AspIleSerPheSerAlaGlyValPheValSerValAla 843
 QY 1346 GCTTATTAATTCATGACCTTTTGGCTTACCTTGACAGTCTTGT---ATACACATTTCTT 1402
 Db 844 AlaThrThrPheCysIleIleAlaLeuTyrAlaLeuThrIlePheLysIleLeuProPhe--- 862
 QY 1403 TTGAGAAAATTCAAATATGATTTTGCAGACGCTTGTGCTTAAATGACTGTATGGGCTAT 1462
 Db 863 -----ThrMetGluIleAspGlnAsnPheIleAlaIleAlaIleLeuAlaIleIleGlyTyr 880
 QY 1463 TCATTAACATATCTTGTGATTCATTTTGTATTCGTATTCGTGAAGATGCCAAGCAACTG 1522
 Db 881 SerLeuAsnAspThrValValValPheAspArgIleArgGluThrMetLysLeuTyrPro 900
 QY 1523 TTTAACCCCTATGCAATGTTTATGATTAATGATCCCTTCAAAAGCGTTACAGCCGAGGTA 1582
 Db 901 AsnArgAspArgTyrGluValIleAsnAspAlaLeuAsnSerThrLeuGlyArgThrLeu 920
 QY 1583 ATGACAACAGCTACACCTCATCATGTTTGTGTAAGCTTTTGTATAGGCGGCTCCTCT 1642
 Db 921 AsnThrSerLeuThrThrPheIleValMetLeuValIlePheIlePheGlyAlaIleThr 940
 QY 1643 GTCTTAAATTTTGCATTTATTAATGACATAGAGATTTCTTACGAACTTATGCTCTT 1702
 Db 941 MetArgSerPheThrPheSerIleLeuLeuGlyIleValIleGlyThrTyrSerThrLeu 960
 QY 1703 TATATTCACACCTCTGTTGTTGTTATGCTCCGTAAGAAATGCTCAAAA 1756
 Db 961 PheValAlaThrPro---LeuAlaTyrGluIleGlnLysArgLysLeuAsnLys 977
 RESULT 5
 AAY34352
 ID AAY34352 standard, Protein, 985 AA.
 AC AAY34352;
 XX 25-AUG-1999 (first entry)
 DT XX
 DE Porphyromonas gingivalis protein pg18.
 XX Porphyromonas gingivalis; Pg, periodontal disease; gingivitis;
 KM vaccine; antigenic.
 XX Porphyromonas gingivalis.
 OS
 PN WO9239870-A1.
 PD 17-JUN-1999.
 PF 10-DEC-1998; 98MD-AU01023.
 XX 04-AUG-1998; 98AU-0005028.
 PR 10-DEC-1997; 97AU-0000839.
 PR 31-DEC-1997; 97AU-0001182.
 PR 30-JAN-1998; 98AU-0001546.
 PR 10-MAR-1998; 98AU-0002264.
 PR 09-APR-1998; 98AU-0002911.
 PR 23-APR-1998; 98AU-0003128.
 PR 05-MAY-1998; 98AU-0003338.
 PR 22-MAY-1998; 98AU-0003654.
 PR 29-JUL-1998; 98AU-0004917.
 XX
 PA (CSLC-) CSL LTD.
 XX Aagus CT, Barr IG, Hocking DM, Margets MB, Patterson MA;
 PI Ross BC, Rochel LJ, Webb EA;

QY 1703 TATATTGCACCACTCTGTTTGTATTAGTCCGTAAGAAAATCGCTCAAAA 1756
 Db 965 PheValAlaIleThrPro---LeuAlaTyGluIleGlnIleArgLysLeuAlaLys 981
 RESULT 6
 AAY03754
 ID AAY03754 standard; Protein; 737 AA.
 XX AAY03754;
 AC AAY03754;
 DT 10-JUN-1999 (first entry)
 DE B. subtilis secretion factor SecDF.
 XX Gram-positive microorganism; secretion factor; hormone; growth factor;
 KW cytolysis; enzyme; protease; cellulase; amylase; carboxylase; lipase;
 KW reductase; isomerase; epimerase; tautomerase; transferase; kinase;
 KW phosphatase; SecDF.
 XX Bacillus subtilis.
 OS Bacillus subtilis.
 PN MO9904007-A1.
 PD 28-JAN-1999.
 XX 16-JUL-1998; 98WO-US14786.
 PF 17-JUL-1997; 97EP-0305344.
 PR 16-JUL-1997; 97EP-0305286.
 XX (GENV) GENENCOR INT. BV.
 PA (GENV) GENENCOR INT. INC.
 XX Quax WJ;
 PI WPI; 1999-132251/11.
 DR N-PSDB; AAX29477.
 XX Expression vector containing sequence for Gram-positive Sec D, F or
 PT PF secretion factors - useful for, e.g. increasing secretion of
 PT heterologous or homologous proteins expressed in host cells
 XX Disclosure; Fig 1A-E; 54pp; English.
 PS The invention relates to an expression vector includes nucleic acid
 CC encoding a Sec D, F or DF secretion factor from a Gram-positive
 CC microorganism, under the control of expression signals that cause
 CC overexpression of the secretion factor in such a microorganism. Cells
 CC containing the vector are used to secrete secretion factors, heterologous
 CC or homologous, that is expressed in host cells. Suitable secretion
 CC factors are hormones, growth factors, cytokines and enzymes (specifically
 CC a protease, cellulase, amylase, carboxylase, lipase, reductase,
 CC isomerase, epimerase, tautomerase, transferase, kinase or phosphatase) or
 CC a ligand, receptor, inhibitor, vaccine or antibody. Expression of these
 CC secretion factors can overcome the bottle neck created by the limited
 CC capacity of the secretory system in Gram-positive bacteria, so increases
 CC production of secreted proteins. The present sequence represents a
 CC Bacillus subtilis secretion factor SecDF.
 XX
 SQ Sequence 737 AA;
 Alignment Scores:
 Pred. No.: 2.5e-67 Length: 737
 Score: 678.50 Matches: 169
 Percent Similarity: 50.43% Conservative: 125
 Best Local Similarity: 28.99% Mismatches: 192
 Query Match: 20.49% Indels: 97
 Gaps: 15
 US-09-868-987-1 (1-1864) x AAY03754 (1-737)
 QY 104 GTCAGCAGCCTATTATTAACGTCCCAATTGAAA---AATCATGCCAGTGTCTCAGGGAAA 160

Db 197 ValSerAlaProAsnValSerGlnGlnIleuAsnThrThrAspValLysIleGlnGlyHis 216
 QY 161 TTTTACCACCGGTAAGTGGCAAACTCGGCTCAGATTTTAAATCTGGAGCGATGCTTTT 220
 Db 217 PheThrAlaGlnGlnAlaLysAspLeuAlaSerIleLeuAsnAlaGlyAlaLeu----- 234
 QY 221 GTTCCCGAGGTTCTCAGTGAAGAG-----ACGATCTCTTGATCTTGGAAAAA 271
 Db 235 ---ProValLysLeuThrGlnLysSerThrSerThrSerValGlyAlaGlnPheGlnGln 253
 QY 272 CAATGTACAAAGGCAATATCTACAGATGCTGTGGCTTGGCAATGTTATTTGATG 331
 Db 254 AlaLeuHisAspThrValPheAlaGlyIleValGlyIleAlaIleIlePheLeuPheMet 273
 QY 332 AGCGTATATTAATGATTGGAGCGGTCATCGGCTTGGAGCGTCTTGTGAATCTTTG 391
 Db 274 LeuPheTyThrArgLeuProGlyLeuIle-----AlaValIleThrLeuSerVal 290
 QY 392 CTATCTGGGAGCTCTACAG-----TATTTGATGCGGCACTCACTTGTGAGA 442
 Db 291 TyrIleTyThrIleThrLeuGlnIlePheAspTyrMetAsnAlaValLeuThrLeuProGly 310
 QY 443 CTCGCTGGGATTTGTTCTTCTATGGGATGCCGATGATGCAATGTTCTGTATTGAA 502
 Db 311 IleAlaIleLeuIleLeuGlyValGlyMetAlaValAspAlaAsnIleIleThrTyGln 330
 QY 503 AGAATCCGAGAGGAATTTTATTTGTCTCAAGCTTTAAAAATCTGTAGAAAAAGATAT 562
 Db 331 ArgIleLeuGlnGlnIleLeuLysLeuGlyLysSerValArgSerValaheArgSerGlyAsn 350
 QY 563 ACCAAGGCTTTGGAGCCATTTTGTGATCTTACCTGATCACTAGATATGGCTTCAGCACT 622
 Db 351 ArgArgSerPheAlaThrIlePheAspAlaAsnIleThrIleIleAlaValAlaVal 370
 QY 623 CTTTCTTCTTACAGACAGGCGCTTAAAGGTTTGTCTTTCATGATTTTGAATTT 682
 Db 371 LeuPheIlePheGlyThrSerSerValLysGlyPheAlaThrMetLeuIleLeuSerIle 390
 QY 683 TTTCTTCAATGTTTACGGCTCTTTTCATGATCACTAAATTTTCTCATGCTG----- 733
 Db 391 LeuThrSerPheIleThrIleValAlaPheLeuSerArgPheLeuAlaLeuLeuValGln 410
 QY 734 -----TGATGATTAAG----- 745
 Db 411 SerArgTyrLeuAspArgLysLysGlyTyrPheGlyValAsnLysLysIleMetAsp 430
 QY 746 -----ACCAACATACACAGTTGCATATGATGAATTAAGTTCTGGGATTAAG 793
 Db 431 IleGlnAspThrAspGlnAsnThrGlnProHisThrProPheGln-----Lys 446
 QY 794 CAGATTTTCTTGAGAGATGCAAAAACTTTGGGCTGTTTGGAGTGTTTTCTTTTA 853
 Db 447 TyrAspPheThrSerLysArgLysTyrPhePheIlePheSerSerValIleThrAla 466
 QY 854 GTTGGCTGCTCGCGGTTTGGAGCTGGAATTCGTTTGGAGATGATTTTAAAGA 913
 Db 467 GlyIleIleIleLeuLeuValPheArgLeuAsn-----LeuGlyIleAspPheAlaSer 484
 QY 914 GGG-----TATGCTTATACCTTAAATCAAAAAGCATGCG 949
 Db 485 GlyAlaArgIleGluValGlnSerAspHisLysLeuThrThrGlnGlnValGlyLysAsp 504
 QY 950 ATCAGCGATTTGCTCAATATGCTGCGCAAGTTGTGCATTAATACAGAGAACTGCTTT 1009
 Db 505 PheGlnSerLeuGlnLysMetAspProAspThrValValLeuSerGlyGlyLysSerAsnIle 524
 QY 1010 TCT---TCTAGAGACTTCGATATCAAACTTGGATCTTCAGAAAAAGATCAAAATGTAT 1066
 Db 525 GlyValAlaArgPheValGlyValProAspLysGlnThrIleAlaLysValLysThrTyr 544
 QY 1067 TTTAGTATTAAGCTTTTAACTTAATCAATCAAGATGAGAGCTCTCTCTTAAATTAACG 1126

Db 545 PheLysAspLys----- 548
QY 1127 ATCATGAGCTGGCTTATTGTGGGATGTTGTACAAAACAGCCCTAGATTCTCTACGGA 1186
Db 549 -----TyrGly 550
QY 1187 AACTCTAACGAAACGCAAAATTTTGGTCAAGGTAAAGCAGCAAACTATCGAAGAAATG 1246
Db 551 SerAspProAsnValSerThrValSerProThrValGlyLysGluLeuAlaArgAsnAla 570
QY 1247 GCTTATCAGGCGACCATCGGCTTTTAGGAGCTTTGGCAATCATCTGTCTCTATGAGT 1306
Db 571 LeuTyrAlaValAlaIleAlaSerIleGly-----IleIleIleTyrValSer 586
QY 1307 TTGCGCTTTGAATGCAATATGCTTTTCAGTGCCTGATGCGCTTTTAATCATGACCTTTTG 1366
Db 587 IleArgPheGluTyrLysMetAlaIleAlaIleAlaSerLeuLeuTyrAsp----- 604
QY 1367 GCTACCTGTCAGTCTGTTTATAGCACATTTCTTTTGAAGAAATTCAAATAGATTG 1426
Db 605 -----AlaPhePheIleValThrPhePheSerIleThrArgLeuGluValAspVal 621
QY 1427 CAAGCATTGGTCTTAAATGACTGATTGGGCTATTCAATAACAATACTTTGATCATT 1486
Db 622 ThrPheIleAlaAlaIleLeuThrIleIleGlyTyrSerIleAsnAspThrIleValThr 641
QY 1487 TTTGATCGTATTCGTGAA-----GATCCCAAGCGAACCTGTTTACCCCTATGCAT 1537
Db 642 PheAspArgValArgGluHisMetLysLysArgLysProLysThrPheAlaAspLeuAsn 661
QY 1538 GTTTAGTTAATGATGCCCTTCAAGACGTTTCAGCCGACCGTAAATGACACACCTACA 1597
Db 662 HisIleValAsnLeuSerLeuGlnGlnThrPheThrArgSerIleAsnThrValLeuThr 681
QY 1598 ACTCTATCAGTTTGTAAATGCTTTTGTATAGCGGCTCCTCTGCTTTTAATTTTGA 1657
Db 682 ValValIleValValThrLeuLeuIlePheGlyAlaSerSerIleThrAsnThrSer 701
QY 1658 TTTATTATGACCATAGGATCTTCTAGGAACCTTTATCGTCTCTTTATATGACACCT 1717
Db 702 IleAlaLeuValGlyLeuLeuThrGlyValTyrSerSerLeuTyrIleAlaGln 721
QY 1718 CTGTTGTTG 1726
Db 722 IleTrpLeu 724
RESULT 7
AAG82042
ID AAG82042 standard; Protein; 741 AA.
XX AAG82042;
XX AC
XX AC
XX DT
XX 03-SEP-2001 (first entry)
XX DE
XX S. epidermidis open reading frame protein sequence SEQ ID NO:1178.
XX DE
XX Staphylococcus epidermidis SR1 strain; infection; diagnosis;
KW vaccination; endocarditis.
XX OS
XX Staphylococcus epidermidis.
XX PN
XX WO200134809-A2.
XX PD
XX 17-MAY-2001.
XX PF
XX 09-NOV-2000; 2000WO-US30782.
XX PR
XX 09-NOV-1999; 99US-0164258.
XX PA
XX (GLAX) GLAXO GROUP LTD.
XX PI
XX Kimmerly WJ;
XX

DR WPI; 2001-316495/33.
DR N-PSDB; AAH52892.
XX Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
PT useful for vaccinating against infections, e.g. endocarditis -
XX
XX Claim 18; Page 338; 2188pp; English.
XX
XX AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides
CC (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis.
CC (I) and (II) can have antibacterial activity and therefore can be used
CC in vaccination. The nucleic acids (I) may be used to produce the
CC S. epidermidis polypeptides (II) via the production of vectors
CC containing them which are used to produce hosts cells which express the
CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
CC used to vaccinate subjects and to raise antibodies against the bacteria.
CC The polypeptides may also be used to assay for other inhibitors of their
CC activity and therefore identify compounds that may be used for the
CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to
CC AAH5090 represent specifically claimed S. epidermidis genomic DNA
CC polynucleotide sequences from the present invention. AAH55091 to
CC AAH55098 represent oligonucleotide sequences and primers which are used
CC in the exemplification of the present invention.
CC N.B. The present invention specifically claims all the polynucleotide
CC sequences given in the sequence listing of the present specification,
CC however the sequence listing only goes up to SEQ ID NO:4454 so even
CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,
CC no sequences are present for SEQ ID NO:4455 to 4464.
XX
XX Sequence 741 AA;
SQ
Alignment Scores:
Pred. No.: 8.25e-64 Length: 741
Score: 647.50 Matches: 165
Percent Similarity: 50.95% Conservative: 129
Best Local Similarity: 28.60% Mismatches: 204
Query Match: 19.55% Indels: 79
DB: 22 Gaps: 12
US-09-868-987-1 (1-1864) x AAG82042 (1-741)
QY 89 ATTACGGTATATGTCAGCAGCGCCTATTTAAACGTCCCATTTGAAATAATCATGCCAGT 148
Db 179 IleSerAlaAlaSerValAspGlnProIleAsnSerSerValGluIleSerGlyGly 198
QY 149 GTCTCAGGGAATTTACCCACCGTGAAGTGAAGCAACTCGCTCAGATTTTAAATCTCGA 208
Db 199 PheAsnGlyLysLysGlyValGluGluAlaLysGlnIleAlaGluLeuAsnAlaGly 218
QY 209 CGCATGTCTTTTGTTCGAGGTTCCTCAGTGAAGAGACGATCTCTCTGTGATCTTGGGAAA 268
Db 219 SerLeuProValAspLeuLysGluIleTyrSerAsnSerValGlyAlaGlnPheGlyGln 238
QY 269 AAACAATGTACAAAGGCAATTATCTCAGCATGCTGTGGCTTGGCAATGCTTATGTTTGG 328
Db 239 AspAlaLeuAspLysThrMetPheAlaSerIleValGlyIleAlaLeuIleTyrLeuPhe 258
QY 329 ATGACGTATATTATAGATTGGAGGCTCATCGCTCCGGAGCTGTCTCTGATCTT 388
Db 259 MetLeuGlyPheTyrArgLeuProGlyLeuValAlaIleIleAlaLeuThrThrTyrIle 278
QY 389 TTGCTTATCTGGCAGCTCTACAGTATTGTGATGGCCACCTCACCTTGTCTCAGGACTCGCT 448
Db 279 TyrLeuThrLeuValAlaPheAsnPheIleSerGlyValLeuThrLeuProGlyLeuAla 298
QY 449 GGGATTGTTCTGTATGCGGATGCGCGTAGATGCAATATGTTCTTGTATTTCGAAGAATC 508
Db 299 AlaLeuValLeuGlyValGlyMetAlaValAspAlaAsnIleIleMetTyrGluArgile 318
QY 509 CGAGAGGAATTTTATTCTCTCAAAGTCTTAAATACTGTAGAAAAGGATATACCAAG 568
Db 319 LysAspGluLeuArgIleGlyArgThrLeuLysGlnAlaTyrSerLysAlaAsnLysSer 338

```

QY 569 GCTTTGAGCCATTTTGAATCTGACTATGAGCTTGGCCAGCACTCTTTTC 628
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 339 SerPheLeuThrIlePheAspSerAsnLeuThrValIleAlaIleValLeuPhe 358
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 629 TTCTGATGATACAGGCGCTTAAAGGGTTCCTTGGACATTTTGAAGATTTTCTC 688
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 359 PhePheGlyGluSerSerValLysGlyPheAlaThrMetLeuLeuGlyIleLeuMet 378
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 689 TCATGTTTACGGCTCTTTTCATGACTTAAATTTTCTTCATGCTG----- 733
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 379 IlePheValThrAlaValAlaPheLeuSerArgGlyLeuLeuSerLeuValSerSerAsn 398
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 734 -----TGATGATATAGAACCCACATACAGCTGATGATGATGAT 775
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 399 PhePheIleYsglnIleTyrIlePheGlyValIleYsglySAspArgHisAspIleAsn 418
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 776 AAGTCGTGGGATTAAGCATGAT-----TTCTTGAGA 808
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 419 GluGlyLysAspVal---HisAspLeuLysThrSerTyrGluArgLeuAsnPheValLys 437
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 809 GGATGCAAAAACCTTGGCGTGTCTTCTGGAAGTCTTTTCTTTAGCTTGGCTGCTCTC 868
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 438 LeuAlaLysProLeuIleSerLeuSerIleLeuIleValIleIleGlyLeuIleIleIle 457
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 869 GGGTTTGAGACCTGGATTCGGTTTGGGATGGATTTTAAAGAGGGTATGCTTACC 928
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 458 SerIlePheLysLeuAsn-----LeuGlyIleAspPheSerSerGlyThrArgAlaAsp 475
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 929 TTTAATCCAAAAGAGATGAGCATGAGCATGCTGCTCAAAATGCGTGGCAAAAGTTGTCAT 988
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 476 IleGlnSerIysAsn-----AlaIleThrGlnAlaGlnValGluLys 489
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 989 AAACCTACAGAGAGCTGCTCTTCTTCTAGAGACTCCGCTATCCAAACATTTGGATCTTCA 1048
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 490 ThrValLysSerValGlyLeuGluProAsp-----GlnIleGlnIleAsnGlySerGly 507
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1049 GAAAGATCAAAATCTATTATTAGATTAAGCTTAAGCATATAGACATGACGAGCC 1108
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 508 AsnLys-----AsnAlaThrValGlnPheLysLys 517
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1109 TCTCTCTAAATTAACGATCATGAGCTGGCGTATTGTGGAGTGTGTGCAGAAACAG 1168
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 518 AspLeuSerArg----- 521
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1169 CCTAGATTCTCTACGGAATCTTAAACGAAACGCAAAATTTGCTCAAG----- 1219
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 522 -----GluGluAspAsnLysLeuSerAlaLysValLysSerGluPheGlyAsp 537
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1220 -----GTAAGACGAAACTATCGAAGAAATATGCGTTATCAGCGCAGCC 1261
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 538 AsnProGlnIleAsnThrValSerProLeuIleGlyGlnLeuAlaLysAsnAlaVal 557
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1262 ATCGGCGCTTTAGAGCTTTGGCAATCATCTGCTCATATGATGATGCTTGGCTTGAATGG 1321
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 558 ThrAlaLeuIleLeuAlaSerIleGlyIleIleIleThrValSerLeuArgPheGluTyr 577
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1322 CAATATCTTACAGTGGCGGTATGCGCTTAAATCATGACTTTTGGCTACTGTGCAGTC 1381
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 578 ArgMetGlyLeuSerSerValLeuAlaLeuLeuHisAspValPhe----- 592
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1382 TTGTTTATAGACATTTCTTTGAAGAAATTCAAATATGATTCGAAGCATTGGTGT 1441
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 593 IleIleIleAlaIlePheSerLeuPheArgLeuGluValAlaPheLeuThrPheIleAlaIle 612
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1442 TTAATGACTGATTTGGGATTCATTAACAATATGATGATGATTTGATGATGATTCGT 1501
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 613 ValLeuThrIleValGlyTyrSerIleAsnAspThrIleValThrPheAspArgValArg 632
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1502 GAAGAT---CGCCAAGCGAAGCTGTTTAC-----CCTATGATGTTTATGTTAAT 1549
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 633 GluAsnLeuHisLysValLysValIleThrHisThrAspGlnIleAspAspIleValAlaAsn 652
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1550 GATGCCCTTCAAAAGACCTTACGCCGACGCTATATGACAAAGCTATCAACTCTATCAGTT 1609
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Db 653 ArgSerIleArgGlnThrThrThrArgSerIleAsnThrValLeuThrValValValVal 672
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1610 TTGTTATACCTTTTGTATTATAGCGGCTCTGCTGCTTAAATTTTGGATTTATGACC 1669
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 673 ValValAlaIleLeuIleLeuGlyAlaProThrIlePheAsnPheSerLeuAlaLeuLeu 692
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1670 ATAGGATTCCTCTAGAACCTTATGCTCTTTATATGACACCACTCTG 1720
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 693 IleGlyLeuLeuSerGlyValPheSerSerIlePheIleAlaValProLeu 709
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 8
ABP38844
ID ABP38844 standard; Protein: 766 AA.
XX
AC ABP38844;
XX
DT 24-JUL-2002 (first entry)
XX
DE Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:3689.
XX
KW Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;
XX antibacterial; gene therapy.
XX
OS Staphylococcus epidermidis.
XX
PN US6380370-B1.
XX
PD 30-APR-2002.
XX
PF 13-AUG-1998; 98US-0134001.
XX
PR 14-AUG-1997; 97US-055779P.
XX PR 08-NOV-1997; 97US-064364P.
XX
PA (GENO-) GENOME THERAPEUTICS CORP.
XX
PI Doucette-Stamm LA, Bush D;
XX
DR WPI, 2002-381255/41.
XX DR N-PSDB; ABN91389.
XX
PT Novel isolated nucleic acid encoding a Staphylococcus epidermidis
XX polypeptide, useful for diagnosing and treating bacterial infections -
XX
PS Disclosure; SEQ ID 3689; 267bp; English.
XX
PX ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading
XX frame (ORF) nucleic acid sequences which encode the amino acid sequences
XX given in ABP35124 to ABP37960. The S. epidermidis sequences have
XX antibacterial activity and can be used in gene therapy. The sequences
XX can also be used in the diagnosis and treatment of bacterial infections,
XX particularly S. epidermidis infections. The sequences can be used to
XX screen for compounds able to interfere with the S. epidermidis life
XX cycle or inhibit S. epidermidis infection.
XX N.B. The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from the
XX USPTO web site.
XX
SQ Sequence 766 AA;

Alignment Scores:
Pred. No.: 8 4e-64 Length: 766
Score: 647.50 Matches: 165
Percent Similarity: 50.95% Conservative: 129
Best Local Similarity: 28.60% Mismatches: 204
Query Match: 19.55% Indels: 79
DB: 23 Gaps: 12

US-09-868-987-1 (1-1864) x ABP38844 (1-766)
QY 89 ATTGACGGTTATATGCTGACGACGCTTATTTTAAACGTCCTCATGAAATCATGCCAGT 148
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Db 204 IleSerAlaAlaSerValAspGlnProIleAsnSerSerValGluIleSerGly 223
 Qy 149 GTCTCAGGGAATTTACCCACGCGTAAGTAGAGCAAACTCGCTCAGATTTAAATCTGGA 208
 Db 224 PheAsnGlyLysLysGlyValGluGluAlaLysGlnIleAlaGluLeuLeuAsnAlaGly 243
 Qy 209 GCGATGCTTTTGTTCGCGAGCTTCTCAGTGAAGAGAGAGATCTCTTCATCTTGGGAA 268
 Db 244 SerLeuProValAspLeuLysGluIleTyrSerAsnSerValGlyAlaGlnPheGlyGln 263
 Qy 269 AAACAATGTACACAGGATATCTCAGCATGCTGCTGGCAATGCTTATCTTTTGG 328
 Db 264 AspAlaLeuAspLysThrMetPheAlaSerIleValGlyIleAlaLeuIleTyrLeuPhe 283
 Qy 329 ATGACGCTATATATAGATTTGGAGCGTCATCGCTTCGGAGCTGTCTCTCGAATCTT 388
 Db 284 MetLeuGlyPheTyrArgLeuProGlyLeuValAlaIleIleAlaLeuThrTyrIle 303
 Qy 389 TTGCTTATCTGGCAGCTCAGATATTGGATGGCCACCTCACCTTCAGGACTCGCT 448
 Db 304 TyrLeuThrLeuValAlaPheAsnPheIleSerGlyValLeuThrLeuProGlyLeuAla 323
 Qy 449 GGGATTGTCTTGTCTATGGGATGGCGCTAGATGCAAAATGTTCTTGATTCGAAGAATC 508
 Db 324 AlaLeuValLeuGlyValGlyMetAlaValAspAlaAsnIleIleMetTyrGluArgIle 343
 Qy 509 CGAGAGGAATTTTATGCTCAAAAGTCTTAAAAATCTGAGAAAAAGATATACCAAG 568
 Db 344 LysAspGluLeuArgIleGlyArgThrLeuLysGlnIleTyrSerLysAlaAsnLysSer 363
 Qy 569 GCTTTGGAGCCATTTTGTATCTAACTGACTACAGTATTGGCTCAGCCTCTCTTTC 628
 Db 364 SerPheLeuThrIlePheAspSerAsnLeuThrValIleAlaAlaValLeuPhe 383
 Qy 629 TTCTCAGATACAGGCCCTATTAAAGGTTTGTCTTGACATTTAGGAATTTCTCT 688
 Db 384 PhePheGlyGluSerSerValLysGlyPheAlaThrMetLeuLeuLeuGlyIleLeuMet 403
 Qy 689 TCAATGTTTACGCTCTTTTCATGACTAAATTTTCTTCATGCTG 733
 Db 404 IlePheValThrAlaValPheLeuSerArgGlyLeuLeuSerLeuValSerSerAsn 423
 Qy 734 -----TGGATGAATAGACCCCAACATACACATGTCATATGATGAT 775
 Db 424 PhePheLysLysGlnTyrTrpLeuPheGlyValLysLysMetAspArgHisAspIleAsn 443
 Qy 776 AAGTTCGTGGGATAAAGCATCAT -----TTCTTGAGA 808
 Db 444 GluGlyLysAspVal--HisAspLeuLysThrSerTyrGluArgLeuAsnPheValLys 462
 Qy 809 GGATGCAAAAACTTTGGGCTGTTTCTGGAAGTGTCTTTTCTTTTAGGTGCGTCTCTC 868
 Db 463 LeuAlaLysProLeuIleSerLeuSerIleLeuIleValIleIleGlyLeuIleIle 482
 Qy 869 GGGTTTGAGCTGGAATCCGTTTGGATGGATTTTAAAGGAGGATGTCCTTTACC 928
 Db 483 SerIlePheLysLeuAsn-----LeuGlyIleAspPheSerSerGlyThrArgAlaAsp 500
 Qy 929 TTTAATCCAAAAGAGCATGGCATCAGCATGTTGTCTCAAAATCGTGCAAGTCTGTCAT 988
 Db 501 IleGlnSerLysAsn-----AlaIleThrGlnAlaGlnValGluLys 514
 Qy 989 AAATCAGGAAGCTGGTCTTTCTTCTAGACATTCCTGATTTCAACATTTGGATCTTCA 1048
 Db 515 ThrValLysSerValGlyLeuGluProAsp-----GlnIleGlnIleAsnGlySerGly 532
 Qy 1049 GAAAGATCAAAATCTATTATTAGTGATAAAGCTTTAAGCTATCTAACAAGATACAGACCC 1108
 Db 533 AsnLys-----AsnAlaThrValGlnPheLysLys 542
 Qy 1109 TCTCTCTAAATTAACGATCATGAGTGGCGCTTATTGTTGGGATGTTGTCAGAACACAGG 1168
 Db 543 AspLeuSerArg----- 546

Qy 1169 CCTAGATTTCTCTACGAAACTCTAAACGAAACCCAAAAATTTTGGTCAAAG----- 1219
 Db 547 -----GluGluAspAsnLysLeuSerAlaLysValLysSerGluPheGlyAsp 562
 Qy 1220 -----GTAAGCAGCAAACTATCGAAGAAATCGCTTATCAGCGCAGCC 1261
 Db 563 AsnProGlnIleAsnThrValSerProLeuIleGlyGlnGluLeuAlaLysAsnAlaVal 582
 Qy 1262 ATCGGGCTTTTAGGAGCTTTGGCAATCATCTGCTCTATGTGAGTTTGGCTTTGAAATGG 1321
 Db 583 ThrAlaLeuIleLeuAlaSerIleGlyIleIleIleTyrValSerLeuArgPheGluTrp 602
 Qy 1322 CAATATGCTTTCAGTGGCTATGCGCTTTAATCATCATGACCTTTGGCTACCTGTCGAGTC 1381
 Db 603 ArgMetGlyLeuSerSerValLeuAlaLeuLeuHisAspValPhe----- 617
 Qy 1382 TTGTTTATAGCACATTTCTTTTGAAGAAATTCAAATAGATTTTGAAGCCATTGGTGCT 1441
 Db 618 IleIleIleAlaIlePheSerLeuPheArgLeuGluValAspLeuThrPheIleAlaAla 637
 Qy 1442 TTAATGACTGTATTGGGCTATTCAATAACAATCTTTGATCATTTTGTATCGTATTTCGT 1501
 Db 638 ValLeuThrIleValGlyTyrSerIleAsnAspThrIleValThrPheAspArgValArg 657
 Qy 1502 GAAGAT---CGCCAAAGCAACCTGTTTACC-----CCTATGCATGTTTATAGTTAAT 1549
 Db 658 GluAsnLeuHisLysValLysValIleThrHisThrAspGlnIleAspAspIleValAsn 677
 Qy 1550 GATCCCTCTCAAAAGAGCTTCAGCCGACGGTATGACACAGCTACACCTCTATCAGTT 1609
 Db 678 ArgSerIleArgGlnThrMetThrArgSerIleAsnThrValLeuThrValValVal 697
 Qy 1610 TTGTTAATGCTTTTGTATTAGCGGCTCTCTCTCTTTAATTTTGCATTTTATATGACC 1669
 Db 698 ValValAlaIleLeuIleLeuGlyAlaProThrIlePheAsnPheSerLeuAlaLeuLeu 717
 Qy 1670 ATAGGATTTCTTAGGAACCTTATCTCTCTCTTTTATATTGCACCCTCTG 1720
 Db 718 IleGlyLeuLeuSerGlyValPheSerSerIlePheIleAlaValProLeu 734
 RESULT 9
 ABB48641
 ID ABB48641 standard; Protein; 754 AA.
 XX ABB48641;
 AC ABB48641;
 XX
 DT 05-FEB-2002 (first entry)
 XX
 DE Listeria monocytogenes protein #1345.
 XX
 KW Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;
 KM vitamin B12; bacterial infection; disease.
 XX
 OS Listeria monocytogenes.
 XX
 PN WO200177335-A2.
 XX
 PD 18-OCT-2001.
 XX
 PF 11-APR-2001; 2001WO-FR01118.
 XX
 PR 11-APR-2000; 2000PR-0004629.
 XX
 PA (INSP) INST PASTEUR.
 XX
 PI Buchrieser C, Frangeul L, Couve E, Rusniok C, Feihl H, Dehoux P;
 PI Dussurget O, Chetouani F, Nedjari H, Glaser P, Kunst F, Cossart P;
 PI Daniels J, Goebel W, Kreft J, Kuhn M, Ng E, Vazquez-Boland JA;
 PI Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A;
 PI Chakraborty T, Domann E, Hain T, Berche P, Charbit A, Durant L;
 PI Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N;
 PI Maduenio E, De Pablos B, Wehland J, Kaerst U, Entian K, Hauf J;

US-09-868-987-1 (1-1864) x AAY37781 (1-126)

Qy 1400 TTTTGAAGAAAATTCAAATAGATTGCAAGCCATTGGTGGCTTTAATCACTGTATTGGG 1459
Db 6 PheCysArgGluLeuGlnIleAspLeuGlnAlaIleGlyAlaLeuMetThrValLeuGly 25
Qy 1460 TATTCATTAACCAATCTTGTATCATTTTGTGATCGTATTCTCGTAGATCGCCAAGCGAAC 1519
Db 26 TyrSerLeuAsnAsnThrLeuIleIlePheAspArgIleAsgGluAspArgGluIle 45
Qy 1520 CTGTTTACCCCTATGCATGTTTGTAGTTAATGATGCCCTTCAAAAGACCGTTCAGCGGCACG 1579
Db 46 LeuPheThrProMetProIleLeuIleAsnAspAlaLeuGlnLysThrLeuGlyArgThr 65
Qy 1580 GTAATGACACAGCTAACTCTACAGTTTGTGTTTAAATGCTTTTGTATAGCGGCTCC 1639
Db 66 ValMetThrThrAlaThrThrLeuSerValLeuValIleLeuLeuPheValGlyGly 85
Qy 1640 TCTGTCCTTAAATTTGTCATTTATTATGACCATAGGATTCCTTAGGAACCTTTATCGTCT 1699
Db 86 SerIlePheAsnPheAlaPheIleMetThrValGlyIleLeuLeuGlyThrLeuSerSer 105
Qy 1700 CTTTATTATGCACCACTCTGTTGTTTGTATGCTCCGTAAAGAAAATCGCTCA 1753
Db 106 LeuTyIleAlaProLeuLeuLeuPheMetValArgLysGluGluAsn 123

RESULT 11
AAW98248
ID AAW98248 standard; Protein; 525 AA.
XX
AC AAW98248;
XX
DT 31-MAR-1999 (first entry)
XX
DE H. pylori GHPO 1127 protein.
XX
KW GHPO protein; Helicobacter infection; gastroduodenal disease; gastrit
XX
KW peptic ulcer disease.
XX
OS Helicobacter pylori.
XX
PN WO9843478-A1.
XX
PD 08-OCT-1998.
XX
PF 01-APR-1998; 98WO-US06371.
XX
PR 29-JUL-1997; 97US-0902615.
XX
PR 01-APR-1997; 97US-0833457.
XX
PR 24-JUN-1997; 97US-0881227.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PA (INMR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.
XX
PI Al-Garawi A, Kleanthous H, Miller C, Oomen RP, Tomb J;
XX
DR WPI; 1998-542293/46.
XX
DR N-PSDB; AAX13967.
XX
PT New isolated Helicobacter polynucleotides - used to develop products
PT for the diagnosis, prevention and treatment of Helicobacter
PT infections and gastrointestinal diseases
XX
XX Claim 8; Page 223-225; 2054pp; English.
XX
CC This sequence represents a Helicobacter pylori GHPO protein of the
CC invention. The polypeptides can be used for preventing or treating
CC Helicobacter infections, and gastroduodenal diseases associated with
CC these infections, including acute, chronic, and atrophic gastritis, an
CC peptic ulcer diseases, e.g. gastric and duodenal ulcers. They can also
CC be used for the production of antibodies. The products can also be used
CC for detection and diagnosis.

Db 314 PheSerValAlaGlnAlaSerAspLeuAlaIleAlaLeuArgSerGlyAlaMetAsnAla 333
 Qy 221 GTTCCGAGTTCTCAGTGAAGAGAGCATCTCTTGATCTCTGGGAAAAACAATGTACA 280
 Db 334 ProIleGlnValLeuGluYbArgIleValGlyProSerLeuGlyLeuYbAspSerIleYb 353
 Qy 281 CAAGCATATCTCAGCATGCTGTGGCTTGGCAATGCTATTGTTTATGAGCGGTATAT 340
 Db 354 ThrSerIleIleAlaLeuValGlyGlyPheIleLeuValMetGlyPheMetAlaLeuYr 373
 Qy 341 TATAGATTGGAGGGCTCATCGCTCGGGAGCTGTCTCTGGAATCTTTTGTATCTGG 400
 Db 374 TyrSerMetAlaGlyValIleAlaCysMetAlaLeuValAlaSerLeuPheLeuIleVal 393
 Qy 401 CGAGCTCTACAGTATTGGATCGCCACTCACCTTGTGAGGACTCGCTGGGATGTTCTT 460
 Db 394 AlaValMetAlaIlePheGlyAlaThrLeuThrLeuProGlyMetAlaGlyIleValLeu 413
 Qy 461 GCTATGGGGATGGCCGTAGATGCAATGTTCTGTATTGTAAGAAATCGGAGAGGAATTT 520
 Db 414 ThrValGlyIleAlaValAspAlaAsnIleIleIleAsnGluArgIleArgGluValLeu 433
 Qy 521 TTATTGTCTCAAGTCTTAAAGTCTGTAGAAAAGGATATACCAAGGCTTTTGGAGCC 580
 Db 434 ArgGluGlyGluValValYbAlaIleHisLeuGlyTyrIleAsnAlaSerArgAla 453
 Qy 581 ATTTTGTGATTCTAAGTCTGACTACAGTATTGGCTCGAGCACTTCTTTCTCTAGATACA 640
 Db 454 IlePheAspSerAsnIleThrSerLeuIleAlaSerValLeuLeuYrAlaTyrGlyThr 473
 Qy 641 GGGCTATTAAAGGTTTGTCTTGACATGATTTTAGGAATTTTCTTCAATGTTTACG 700
 Db 474 GlyAlaIleYbGlyPheAlaLeuThrThrGlyIleGlyIleLeuAlaSerIleIleThr 493
 Qy 701 GCTCTTTTTCATGACTAAATTTTCTTCATGCTGTGATGAATAAGCAACCAATACACAG 760
 Db 494 AlaIleIleGlyThrGlnGlyIleTyrGlnAlaLeuLeuProLYsLeuAlaGlnThrLYs 513

RESULT 13
 AAB88551
 ID AAB88551 standard; Protein; 616 AA.
 XX
 AC AAB88551;
 XX
 DT 04-JUN-2001 (first entry)
 XX
 DE Haemophilus influenzae essential bacterial protein SEQ ID NO:120.
 XX
 KW Haemophilus influenzae; essential bacterial gene; identification;
 KW otitis media; meningitis; upper respiratory tract infection;
 KW infection; antimicrobial.
 XX
 OS Haemophilus influenzae.
 XX
 FN WO200111033-A2.
 XX
 PD 15-FEB-2001.
 XX
 PF 03-AUG-2000; 2000WO-US21176.
 XX
 PR 04-AUG-1999; 99US-0368382.
 XX
 PA (ABBO) ABBOTT LAB.
 XX
 PI Chovan LE, Hessler PE, Reich KA;
 XX
 DR WPI; 2001-147511/15.
 XX
 DR N-ESDB; AAF94404.
 XX
 PT Essential bacterial genes from Haemophilus influenzae and methods for
 PT identifying 'essential' genes that may be potential therapeutic targets
 XX

PS Claim 9; Page 172-173; 185pp; English.
 XX
 CC AAF94345 to AAF94409 represent essential bacterial genes from
 CC Haemophilus influenzae, which encode the proteins given in AAB88492 to
 CC AAB88556. The present invention also describes methods for identifying
 CC essential bacterial genes (i.e. those essential to the survival of a
 CC bacterium) using a transposition system. The methods are used to
 CC identify essential genes from bacteria, especially H. influenzae (which
 CC causes otitis media, meningitis and upper respiratory tract infections)
 CC which may be used as targets for potential antimicrobial agents.
 CC AAF94410 to AAF94416 represent PCR primers used in the exemplification
 CC of the present invention.
 XX
 SQ Sequence 616 AA;
 Alignment Scores:
 Pred. No.: 9,62e-37 Length: 616
 Score: 408.50 Matches: 87
 Percent Similarity: 65.85% Conservative: 48
 Best Local Similarity: 42.44% Mismatches: 59
 Query Match: 12.33% Indels: 11
 DB: Gaps: 2
 US-09-868-987-1 (1-1864) x AAB88551 (1-616)
 Qy 137 AATCATGCCAGTGTCTCAGGGAATTT-----ACC 166
 Db 395 AsnValAlaThrIleGlnGlyArgPheGlySerAsnPheGlnIleThrGlyValAspSer 414
 Qy 167 CACGCTGAAGTGACCAACTCGCTCAGATTAAATCTGGAGCGATGCTTTTGTTCCTC 226
 Db 415 IleAlaGluAlaHisAsnLeuSerThrLeuLeuYbSerGlyAlaLeuAlaProIle 434
 Qy 227 GAGGTCTCAGTGAAGAGAGCATCTCTTCGATCTCGGAAAAACAATGTACACAAGGC 286
 Db 435 GlnIleValGluGluArgThrIleGlyProSerLeuGlyAlaGlnAsnValGluGlnGly 454
 Qy 287 ATTATCTCAGCATCTCTGGCTTGGCAATGCTATTGTTTTCAGCGGTATATTATAGA 346
 Db 455 IleAsnAlaSerLeuTrpGlyLeuValAlaValIleAlaPheMetLeuPheTyrTyrLYs 474
 Qy 347 TTTGGAGCGTCATCGCTTCGGAGCTGTTCTCTGTAATCTTTTGTCTATCTGGCGAGCT 406
 Db 475 MetPheGlyValIleAlaSerPheAlaLeuValIleAsnIleValLeuValGlyLeu 494
 Qy 407 CTACAGTATTG---GATGGCCACTCACCTTGTGAGGACTCGCTGGGATGTTCTTCT 463
 Db 495 MetSerIleLeuProGlyAlaThrLeuSerMetProGlyIleAlaGlyIleValLeuThr 514
 Qy 464 ATGGGGATGGCGGTAGATGCAAAATGTTCTGTATTGAAAGAAATCCGAGAGGAATTTTA 523
 Db 515 LeuGlyMetSerValAspAlaAsnValLeuIlePheGluArgIleLYeGluGluIleArg 534
 Qy 524 TTGTCTCAAGTCTTAAATAATCTGTAGAAAAGATATACCAAGGCTTTTGGAGCCATT 583
 Db 535 AsnGlyArgSerIleGlnAlaIleAsnGluGlyTyrAsnGlyAlaPheThrSerIle 554
 Qy 584 TTTGATTTCTAACTTCACTACAGTATTGCGCTCACGACTTCTTTTCTCTAGATACAGGG 643
 Db 555 PheAspAlaAsnLeuThrThrIleLeuThrAlaIleLeuLeuTyrAlaValGlyThrGly 574
 Qy 644 CCTATTAAAGGGTTTGTCTTGACATTGATTTTAGGAATTTTCTCTCAATGTTTACGGCT 703
 Db 575 ProIleGlnGlyPheAlaIleThrLeuSerLeuGlyValAlaIleSerMetPheThrAla 594
 Qy 704 CTTTTCATGACTAAA 718
 Db 595 IleThrGlyThrArg 599
 RESULT 14
 AAU91482
 ID AAU91482 standard; Protein; 616 AA.
 XX

AC AAU91482;
 XX
 DT 18-JUN-2002 (first entry)
 XX
 DE Haemophilus influenzae essential gene #60.
 XX
 KW Essential bacterial gene; antifungal agent; antibacterial agent;
 KW antiparasitic agent; insecticidal agent; microbial infection;
 KW mucous membrane infection; otitis media; sinusitis; bronchitis;
 KW alveolitis; conjunctivitis; pneumonia; meningitis; epiglottitis;
 XX cellulitis; septic arthritis.
 OS Haemophilus influenzae.
 XX
 PN WO200218601-A2.
 XX
 PD 07-MAR-2002.
 XX
 PF 22-AUG-2001; 2001WO-US26245.
 XX
 PR 25-AUG-2000; 2000US-0649145.
 XX
 PA (ABBO) ABBOTT LAB.
 XX
 PI Chovan LE, Heesler PE, Reich KA;
 XX
 DR MPI; 2002-304258/34.
 DR N-PSDB; ABR64968.
 XX
 PT Essential bacterial genes in Haemophilus influenzae necessary for
 PT bacterium's growth and survival, useful for screening inhibitors of
 PT polypeptides and developing therapeutic agents e.g. antimicrobial
 XX
 PS Claim 9; Page 172-173; 185pp; English.
 XX
 CC The invention describes an essential bacterial gene (I) comprising a
 CC purified polynucleotide isolated from Haemophilus influenzae where (I)
 CC is essential to H. influenzae survival. The encoded polypeptide (II) is
 CC useful for screening substances that function to inhibit essential H.
 CC influenzae polypeptides by contacting (II) with the desired substances
 CC and measuring the response by a screen from specific, enzyme, general,
 CC affinity, phenotypic and binding screen. (I) and (II) are useful in
 CC developing therapeutic agents such as antifungal, antibacterial and
 CC antiparasitic agent, insecticidal agent, and preventive antimicrobial
 CC agents which are effective in preventing microbial infection or useful
 CC in treatment of that particular infection. (I) and (II) may also be
 CC useful in treatment of mucous membrane infections such as otitis media,
 CC sinusitis, bronchitis, alveolitis, conjunctivitis, pneumonia, meningitis,
 CC epiglottitis, cellulitis and septic arthritis. This is the amino acid
 CC sequence of an essential H. influenzae gene, described in the invention.
 XX
 SQ Sequence 616 AA;
 Alignment Scores:
 Pred. No.: 9.62e-37 Length: 616
 Score: 408.50 Matches: 87
 Percent Similarity: 65.85% Conservative: 48
 Best Local Similarity: 42.44% Mismatches: 59
 Query Match: 12.33% Indels: 11
 DB: 23 Gaps: 2
 US-09-868-987-1 (1-1864) x AAU91482 (1-616)
 QY 137 AATCATGCCAGTGTCTCAGAGAAATT-----ACC 166
 DB 395 AsnValAlaThrIleGlnGlyArgPheGlySerAsnPheGlnIleThrGlyValAspSer 414
 QY 167 CACCGTAGAGCAAACTGCGCTCAGATTAAATCTGGAGCGATGCTTTGTCC 226
 DB 415 IleAlaGluAlaHisAsnLeuSerThrLeuLeuYSerGlyAlaLeuIleAlaProIle 434
 QY 227 GAGGTTCTCACTGAAGAGCATGCTTTCATCTTGGAAGAAAACAATGACCAAGGC 286
 DB 595 IleThrGlyThrArg 599

DB 435 GlnIleValGluGluArgThrIleGlyProSerLeuGlyAlaGlnAsnValGluGlnGly 454
 QY 287 ATTATCTCAGCATGCTGTGGCTTGGCAATGCTATTGTTTGGATGCGGTATATTAAGA 346
 DB 455 IleAsnAlaSerLeuThrPrgIleValAlaValIleAlaPheMetLeuPheTyrTyrLys 474
 QY 347 TTGGAGGGGCTCATTCGCTTCGGAGCTGTTCTTCGAATCTTTCGTTATTCGGACGCT 406
 DB 475 MetPheGlyValIleAlaSerPheAlaLeuValIleAsnIleValLeuValGlyLeu 494
 QY CTACAGATTTTG--GATGGCCACTCACCTTGTCAGACCTCGCTGGAGATTGTTTGCT 463
 DB 495 MetSerIleLeuPrgIylAlaThrLeuSerMetPrgIylIleAlaGlyIleValLeuThr 514
 QY 464 ATGGGAGATGGCCGTAGATGCAATGTTCTTGTATTTGAAAGAAATCCGAGAGAAATTTTA 523
 DB 515 LeuGlyMetSerValAspIleAsnValIleIlePheGluArgIleGlyGluIleArg 534
 QY 524 TTGTCGCAAGCTTMAAAATCTGTAGAAAAGATATACCAAGCTTTGGAGCCATT 583
 DB 535 AsnGlyArgSerIleGlnGlnAlaIleAsnGlyTyrAsnGlyAlaPheThrSerIle 554
 QY 584 TTGATTTCTACTTGACTGACTACGATATGGCCTCAGACCTTCTTCTCTAGATACAGGG 643
 DB 555 PheAspAlaAsnLeuThrIleGlnAlaIleAlaIleLeuTyrAlaValGlyThrGly 574
 QY 644 CTTATTAAGGGTTTCTTGGACATTTGAAATTTTCTTCAATGTTTACGGCT 703
 DB 575 ProIleGlnGlyPheAlaIleThrLeuSerLeuGlyValAlaIleSerMetPheThrAla 594
 QY 704 CTTTTCATGACTAAA 718
 DB 595 IleThrGlyThrArg 599
 RESULT 15
 AAU91069
 ID AAU91069 standard; Protein; 618 AA.
 XX
 AC AAU91069;
 XX
 DT 05-JUN-2002 (first entry)
 XX
 DE Neisseria cell surface polypeptide #7.
 XX
 KW Cell surface protein; antibacterial; antimicrobial.
 XX
 OS Neisseria meningitidis.
 XX
 PN WO200216612-A2.
 XX
 PD 28-FEB-2002.
 XX
 PF 21-AUG-2001; 2001WO-GB03759.
 XX
 PR 24-AUG-2000; 2000GB-0020952.
 XX
 PA (MICS-) MICROSCIENCE LTD.
 XX
 PI Lane JD, Hughes MJG, Santangelo JD;
 XX
 DR MPI; 2002-280941/32.
 DR N-PSDB; ABR54085.
 XX
 PT Novel peptide encoded by Neisseria meningitidis, useful for manufacture
 PT of medicament for treatment or prevention of condition associated with
 PT infection by Neisseria or Gram-negative bacteria -
 XX
 PS Claim 5; Page 44-46; 79pp; English.
 XX
 CC The invention relates to polypeptides located on the cell surface of
 CC Neisseria meningitidis, and the polynucleotides encoding them. The
 CC sequences of the invention are useful for therapeutic or diagnostic use,
 CC in the manufacture of a medicament for use in treatment or prevention of


```
QY 104 GTCAGCAGCCCTATTAAATGCCATTTGAAAAATCATGCCAGTGTCTAGGAAATTT 163
Db 551 ILeSnIleAlaenIIeGlnSerArgLeuGlyAsnSerPheAlaGlyIleHnGlyIleAsn 570
QY 164 ACCCAGCCGTGAAGTACGAAATCGCCCTCAGATTAAATTCAGGAGCATCTTTTGT 223
Db 571 AsnProAsnGlnAlaArgIleuSerIleuLeuArgIleAlaIleuIleAlaPro 590
QY 224 CCCGAGGTTCTCAGTGAAGACATCTCTTGTATCTTGGGAAAAACAATGATACAA 283
Db 591 ILeGlnIleValGlnGluArgThrIleGlyProThrLeuGlyMetGlnAsnIleGln 610
QY 284 GGCATTCTCAGCATGTGTGGGCAATGCTTATGTTTGAAGAGCGATATATAT 343
Db 611 GlyLeuGlnAlaCysLeuAlaGlyLeuLeuValSerIleLeuPheMetIleIlePhe 630
QY 344 AGATTGGAGCGCTCATCGCTTCGAGAGCTGTTCTTGAATCTTTTGTCTTATCTGGGCA 403
Db 631 LysIlePheGlyLeuIleAlaThrSerAlaLeuIleAlaAsnLeuIleValGly 650
QY 404 GCTCTACAGTATTTG--GATGCCCACTCACCTTTCAGAGACTCGCTGGATTTGTT 460
Db 651 ILeMetSerLeuLeuProGlyAlaThrLeuSerMetProGlyIleAlaGlyIleValLeu 670
QY 461 GCTATGGGAGTGGCCGTAGATGCAATGTTCTTGAATTCGAAAGAATCCGAGAGAAAT 520
Db 671 ThrLeuAlaValAlaValAspAlaAsnValLeuIleAsnGluValGlyLeuGlnLeu 690
QY 521 TTATTGCTCAAACTTTAAATCTGTAGAAAAGATATACCAAGGCTTTTGGAGCC 580
Db 691 SerAsnGlyArgThrValGlnGlnAlaIleAspGlnGlyTrpArgGlyAlaPheSerSer 710
QY 581 ATTTGATTTAACTGATGATCAAGTATGGCTCAGACCTTCTTCTTCTTCTGATACAA 640
Db 711 ILePheAspAlaAsnIleThrThrLeuIleLeuValIleIleLeuValAlaValGlyThr 730
QY 641 GGGCTTAAAGGTTGCTTGAACATTTGATTTAGGAATTTCTCTCAATGTTTACG 700
Db 731 GlyAlaIleLeuGlyPheAlaIleThrThrGlyIleGlyValAlaIleAsnSerMetPheThr 750
QY 701 GCTCTTTTCATGACTAAA 718
Db 751 AlaIleValAlaGlyThrArg 756
RESULT 17
ABG26197
ID ABG26197 standard; Protein; 713 AA.
XX
AC ABG26197;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic p:protein #26188.
XX
KM Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
XX
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI, 2001-639362/73.
```

```
DR N-PSDB; AAS90384.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity
XX
PS Claim 20; SEQ ID No 56556; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 713 AA;
Alignment Scores:
Pred. No.: 6,34e-31 Length: 713
Score: 357.50 Matches: 75
Percent Similarity: 65.48% Conservative: 54
Best Local Similarity: 38.07% Mismatches: 67
Query Match: 10,79% Indels: 1
DB: Gaps: 1
US-09-868-987-1 (1-1864) x ABG26197 (1-713)
QY 131 TTGAAAAATATGCCAGTGTCTCAGGAAATTTACCCAGCGTAGAGCAATCGCC 190
Db 500 LeuGlyAsnSerPheArgIleHnGlyIleAsnAsnProAsnGlnAlaArgGlnLeuSer 519
QY 191 TCAGATTAAATCTGAGCGATGTCTTGTTCGCCGAGTTCACGTGAAAGACATC 250
Db 520 LeuLeuLeuArgAlaGlyAlaLeuIleAlaProIleGlnIleValGlnGluArgThrIle 539
QY 251 TCTTGTGATCTGGAAAAACAATGATACACAGCATTTATCTCAGCATGCTGCTTG 310
Db 540 GlyProThrLeuGlyMetGlnAsnIleGlnGlyLeuGlnAlaCysLeuAlaGlyLeu 559
QY 311 GCATGCTTATTTTGTGAGCGTATATATAGATTGGAGCGCATCGCTGGGGA 370
Db 560 LeuValSerIleLeuPheMetIleIlePheTrpLysPheGlyLeuIleAlaIleSer 579
QY 371 GCTGTTCTTGAATCTTTTGTGATCTGGGAGCTCTACAGTATTTG--GATGGCGCA 427
Db 580 AlaLeuIleAlaAsnLeuIleLeuIleValGlyIleMetSerLeuLeuProGlyAlaThr 599
QY 428 CTCACCTTGTACGACTCGCTGGATTTGTTCTTATGCGGATGCGCGTATGCAAT 487
Db 600 LeuSerMetProGlyIleAlaGlyIleValLeuThrLeuAlaValAlaPheAlaAsn 619
QY 488 GTTCTGTATTCGAAAGAACCCAGAGCAATTTTATGCTCAAGCTCTTAAATATCT 547
Db 620 ValLeuIleAsnGlnArgIleGlyGlnGlnLeuSerAsnGlyArgThrValGlnAla 639
QY 548 GTAGAAAAAGATATACCAAGCTTTTGGAGCATTTTGTATCTTAACTGATCAGTA 607
```

Db 640 IleaspGluGlyTyrArgGlyAlaPheSerSerIlePheAspAlaAsnIleThrThrLeu 659

Qy 608 TTGGCTCAGCATTCTTTCTCTAGATACAGGGCCTATTAAAGGTTGCTTTGACA 667
 ::::::::::: ::::

Db 660 IleLysValIleIleLeuTyrAlaValGlyThrGlyAlaIleLysGlyPheAlaIleThr 679
 ::::::::::: ::::

Qy 668 TTGATTTAGGAATTTCTCTCAATGTTTACGGCTCTTTTCATGACATAAA 718
 ::::::::::: ::::

Db 680 ThrGlyIleGlyValAlaThrSerMetPheThrAlaIleValGlyThrArg 696

RESULT 18

AAW98768

ID AAW98768 standard; Protein; 323 AA.

XX

AC AAW98768;

XX

DT 31-MAR-1999 (first entry)

XX

DE H. pylori GHPO 1125 protein.

XX

KW GHPO protein; Helicobacter infection; gastroduodenal disease; gastritis;
 peptic ulcer disease.

XX

OS Helicobacter pylori.

XX

FN WO9843478-A1.

XX

PD 08-OCT-1998.

XX

PF 01-APR-1998; 98WO-US06371.

XX

PR 29-JUL-1997; 97US-0902615.

PR 01-APR-1997; 97US-0833457.

PR 24-JUN-1997; 97US-0881227.

XX

PA (HUMA-) HUMAN GENOME SCI INC.

FA (INWR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.

XX

FI Al-Garawi A, Kleanthous H, Miller C, Oomen RP, Tomb J;
 N-PSDB; AAX11487.

XX

DR WPI; 1998-542293/46.

DR N-PSDB; AAX11487.

XX

PT New isolated Helicobacter polynucleotides - used to develop products
 for the diagnosis, prevention and treatment of Helicobacter
 infections and gastrointestinal diseases

PT

XX

PS Claim 8; Page 1658-1659; 2054pp; English.

XX

CC This sequence represents a Helicobacter pylori GHPO protein of the
 invention. The polypeptides can be used for preventing or treating
 Helicobacter infections, and gastroduodenal diseases associated with
 these infections, including acute, chronic, and atrophic gastritis, and
 peptic ulcer diseases, e.g. gastric and duodenal ulcers. They can also be
 used for the production of antibodies. The products can also be used for
 detection and diagnosis.

XX

SQ Sequence 323 AA;

Alignment Scores:

Pred. NO.:	2.26e-30	Length:	323
Score:	351.00	Matches:	91
Percent Similarity:	50.81%	Conservative:	66
Best Local Similarity:	29.45%	Mismatches:	104
Query Match:	10.60%	Indels:	48
DB:	19	Gaps:	7

US-09-868-987-1 (1-1864) x AAW98768 (1-323)

Qy 830 GTTTCGGAAGTGTCTTTCTTTAGTTGGTTCCTCGGGTTTGAGCCCTGGAAATCC 889
 ::::::::::: ::::

Db 22 ValSerAlaIleLeuAlaLeuAlaLeuAlaLeuGlyLeuPhePheGlyPheSer--- 40

Qy 890 GTTTTGGGAATGGATTTTAAAGGAGGGTATGCCCTTTACCTTTTAAATCCAAAAGACGATGCC 949
 ::::::::::: ::::

Db 41 ---LeuGlyIleAspPheAlaGlyGly----- 48

Qy 950 ATCAGCGATGTGCTCAAAATGCGGGC-----AAAGTTGTGCATAAA 991
 ::::::::::: ::::

Db 49 -----SerLeuValGlnValArgTyrThrGlnAsnAlaProIleLysGluValArgAsp 66

Qy 992 CTACAGGAAGCTGTCTTTCTTCTAGAGACTTCGGTATTCAACACATTTGGATCTTCAGAA 1051
 ::::::::::: ::::

Db 67 LeuPheGluLysGluAlaArgPheLysGlyValGlnValSerGluPheGlySerLysGlu 86

Qy 1052 AAGATCAAAATCTATTATTAGTATAAAGCTTTAAGCTATTACTAAGCAGATACGAGCCCTCT 1111
 ::::::::::: ::::

Db 87 GluIleLeuIleLysPheProPheValGluThrAlaGluAsnGluAspLeuAsnAla--- 105

Qy 1112 CTCCTAAATTAACGATCATGAGCTGGCGTTATTGTGGGATTGTGTGCAGAAACAGGCCT 1171
 ::::::::::: ::::

Db 106 -----IleValAlaAsn----- 109

Qy 1172 AGATTTCTCTACGGAAACTCTTAAACGAAACGCAAAATTTTGGTCAAAAGGTAAAGCAGCAA 1231
 ::::::::::: ::::

Db 110 ---IleLeuLysProSerGlyAspPheGluIleArgLysPheAspThrValGlyProArg 128

Qy 1232 CTATCGAAGAAATGCGTTATCAGCGCACCATCGGCTTTTAGGAGCTTTGGCAATCATC 1291
 ::::::::::: ::::

Db 129 ValGlySerGluLeuLysGlyIleLeuSerLeuIleAlaIleAlaIle 148

Qy 1292 TTGCTCTATGTAGTTTGGCTTTGAATGCGCAATATGCTTTCAGTGGCGTATGGCTTTA 1351
 ::::::::::: ::::

Db 149 MetValTyrValSerPheArgTyrGluTrpArgPheAlaLeuAlaSerValIleAlaLeu 168

Qy 1352 ATTCATGAC-----CTTTTGGCTACCTGTGCGACTCTTTTATATAGCACATTTCTTTTG 1405
 ::::::::::: ::::

Db 169 ValHisaspValIleLeuValAlaSerSerValIleValPhe----- 182

Qy 1406 AAGAAATTCAAATAGATTGCAAGCCATTGGTCTTTAATGACTGTATGGGGTATTCA 1465
 ::::::::::: ::::

Db 183 ---LysIleAspMetAsnLeuGluValIleAlaAlaLeuLeuThrLeuIleGlyTyrSer 201

Qy 1466 TTAACAATATCTTTGATCATTTTTCATCGTATTCGTGAGATCCCAAGCGCAACCTGTTT 1525
 ::::::::::: ::::

Db 202 IleAsnAspThrIleIlePheAspArgIleArgGluGluMet***SerGlnLysThr 221

Qy 1526 ACCCTATGATGTTTGTAGTTAATAGCCCTTCAAAAGACGTTTCAGCGCAGCGTAATG 1585
 ::::::::::: ::::

Db 222 LysAsnAlaThrGlnAlaIleAspGluAlaIleSerSerThrLeuThrArgThrLeuLeu 241

Qy 1586 ACAACAGCTACAACCTCTATCAGTTTGTATATGCTTTTGTATAGCGGCTCTCTCTGTC 1645
 ::::::::::: ::::

Db 242 ThrSerLeuThrValPhePheValValIleLeuLeuCysValPheGlySerLysIleIle 261

Qy 1646 TTTAATTTTGCATTTATATGACCATAGGATTTCTTCTAGGAACCTTTATCTCTCTTTAT 1705
 ::::::::::: ::::

Db 262 IleGlyPheSerLeuProMetLeuIleGlyThrIleValGlyThrThrSerSerIlePhe 281

Qy 1706 ATTGCACACCTCTGTTGTTGTTATG 1732

Db 282 IleAlaProLysValAlaLeuLeuLeu 290

RESULT 19

AAW37782

ID AAW37782 standard; Protein; 140 AA.

XX

AC AAW37782;

XX

DT 07-OCT-1999 (first entry)

XX

DE Amino acid sequence of a Chlamydia trachomatis protein.

XX

KW Vaccine; eye disease; conventional trachoma; nonendemic trachoma;
 paratrachoma; inclusion conjunctivitis; genital disease; perihepatitis;
 nongonococcal urethritis; epididymitis; cervicitis; salpingitis;

KW

KW bartholinitis; pneumopathy; venereal lymphogranulomatosis.
 XX Chlamydia trachomatis.
 OS
 PN MO9928475-A2.
 XX
 PD 10-JUN-1999.
 XX
 PF 27-NOV-1998; 98WO-IB01939.
 XX
 PR 04-NOV-1998; 98US-0107077.
 PR 28-NOV-1997; 97FR-0015041.
 PR 17-DEC-1997; 97FR-0016034.
 XX
 PA (GEST) GENSET.
 XX
 PI Griffais R;
 XX
 DR WPI, 1999-371125/31.
 XX
 PT Genome sequence of Chlamydia trachomatis
 XX
 PS Disclosure; Page 1368; 1755pp; English.
 XX
 CC AAY36754-Y37949 are encoded by open reading frames (ORFs) of the genome
 CC of Chlamydia trachomatis (see AA201425). The polypeptides can be used as
 CC vaccines against Chlamydia trachomatis. Antisense and ribozyme sequences
 CC can also be used to control growth of the microorganism. Chlamydia
 CC trachomatis is responsible for a large number of diseases, e.g. eye
 CC diseases such as conventional trachoma, nonendemic trachoma,
 CC paratrachoma, and inclusion conjunctivitis; genital diseases such as
 CC nongonococcal urethritis, epididymitis, cervicitis, salpingitis,
 CC peritphatits, bartholinitis; pneumopathy in breast feeding infants;
 CC and venereal lymphogranulomatosis. The polypeptides of the invention
 CC may be of use in treating these diseases.
 XX
 SQ Sequence 140 AA;
 Alignment Scores:
 Pred. No.: 2,43e-30 Length: 140
 Score: 349.00 Matches: 82
 Percent Similarity: 69.12% Conservative: 12
 Best Local Similarity: 60.29% Mismatches: 32
 Query Match: 10.92% Indels: 11
 DB: 20 Gaps: 1
 US-09-868-987-1 (1-1864) x AAY37782 (1-140)
 QY 1770 TTTACGGTACTTATTTTGGCGATT-----TCT 1741
 Db 5 PheArgIleThrTyrAlaGlnAlaTyrPheIleIleTyrArgArgGluPheCysSerSer 24
 QY 1740 TTAGGACCATTAACACACAGAGGTCGTCATTAATTAAGACGATTAAGTCTCTAGA 1681
 Db 25 LeuArgThrMetCysAsnLysArgSerGlyGlyAlaMetCysTranGluLysSerValProAsn 44
 QY 1680 AGAATCCCTATGTCATTAATTAATGCAAAATTAAGACAGAGGACCGCCCTTAATTAACAAA 1621
 Db 45 LysIleProThrValIleMetCysAsnAlaLysLeuLysIleGluProProPheThrAsnAsn 64
 QY 1620 AGCATTAACAAACTGATAGAGTTGAGCTGTGTGTCATTACCGTGGCGCTGAAGCTCTT 1561
 Db 65 LysMetThrSerThrAspArgValAlaValAlaValIleThrValArgProAsnValPhe 84
 QY 1560 TGAAGGCGATCATTAACATAACATGCATAGGGTAAACAGATTGCGTGGCGATCTTCA 1501
 Db 85 CysAsnAlaSerLeuIleLysIleGlyIleGlyValAsnSerPheSerArgArgSerSer 104
 QY 1500 CGAATACGATCAAAATGATCAAAATGATTTTATGATGATACCCCAATATACGATCAAAA 1441
 Db 105 ArgIleArgSerLysIleIleAsnValLeuPheLysGluTyrProSerThrValIleAsn 124
 QY 1440 GCACCAATGGCTGCATTAATTTTG-AAATTTCTTCAAAAAGAAATG 1394

Db 125 AAlProIleAlaCysLysSerIleCysAsnSerLeuGlnLysGluMet 140
 RESULT 20
 AAB88556
 ID AAB88556 standard; Protein; 325 AA.
 XX
 AC AAB88556;
 XX
 DT 04-JUN-2001 (first entry)
 XX
 DE Haemophilus influenzae essential bacterial protein SEQ ID NO:130.
 XX
 KW Haemophilus influenzae; essential bacterial gene; identification;
 KW otitis media; meningitis; upper respiratory tract infection;
 KW infection; antimicrobial.
 XX
 OS Haemophilus influenzae.
 XX
 PN WO200111033-A2.
 XX
 PD 15-FEB-2001.
 XX
 PF 03-AUG-2000; 2000WO-US21176.
 XX
 PR 04-AUG-1999; 99US-0368382.
 XX
 PA (ABBO) ABBOTT LAB.
 XX
 PI Chovan LE, Hessler PE, Reich KA;
 XX
 DR WPI, 2001-147511/15.
 DR N-PsDB; AAF94409.
 XX
 PT Essential bacterial genes from Haemophilus influenzae and methods for
 PT identifying 'essential' genes that may be potential therapeutic targets
 PT
 PS Claim 9; Page 183-184; 185pp; English.
 XX
 CC AAF94345 to AAF94409 represent essential bacterial genes from
 CC Haemophilus influenzae, which encode the proteins given in AAB88492 to
 CC AAB88556. The present invention also describes methods for identifying
 CC essential bacterial genes (i.e. those essential to the survival of a
 CC bacterium) using a transposition system. The methods are used to
 CC identify essential genes from bacteria, especially H. influenzae (which
 CC causes otitis media, meningitis and upper respiratory tract infections)
 CC which may be used as targets for potential antimicrobial agents.
 CC AAF94410 to AAF94416 represent PCR primers used in the exemplification
 CC of the present invention.
 XX
 SQ Sequence 325 AA;
 Alignment Scores:
 Pred. No.: 2,38e-29 Length: 325
 Score: 342.00 Matches: 98
 Percent Similarity: 51.90% Conservative: 80
 Best Local Similarity: 28.57% Mismatches: 11
 Query Match: 10.33% Indels: 54
 DB: 22 Gaps: 15
 US-09-868-987-1 (1-1864) x AAB88556 (1-325)
 QY 710 ATGACTAAATTTTCTTCATGCTGTGGATGAATTAAGACCCACATACACAGTTGATATG 769
 Db 1 MetMetLysLeuHe-----ThrLysAspLysAspGlyHisPhe 13
 QY 770 ATGAATTAAGTTCGTGGGATTAAG-----CATATTTCTTGAAGAGATGC 814
 Db 14 IleArgGluIleAsnGlyIleLysLeuProPheProLeuThrGluPheMetLysValArg 33
 QY 815 AAAAATTTGGGCGTGTTCGTAAGAGTGTTCCTTTAGCT-----TCGGTTGCT 865

Db 34 LysLeuGlyTyrIleLeuSerAlaLeuLeuMetValIleSerLeuPhePheIleIleThr 53
 Qy 866 CTCGGTTGGAGCCTCGAATCCGTTTGGGAATGGAATTTAAAGAGGATGCGCTTT 925
 Db 54 LysGlyPheAsn---Trp-----GlyLeuAspPheThrGlyGlyValPhe 68
 Qy 926 ACCTTTAAATCCAAAGACATGGCATCAGCGATGTGCTCAATCGTGGCAAGTTGTG 985
 Db 69 AspThrHisPheSerGlnSer-----AlaAsnLeuGluGlnIleArgSer----- 83
 Qy 986 CATAAATACAGGAGCTGGTCTTCTCTAGAGACTTCCGTTATTCAAAACATTTGGATCT 1045
 Db 84 ---LysLeuHisGluAsnGlyIleGluSerPro-----IleValGlnThrThrGlySer 100
 Qy 1046 TCAGAAAGATCAAAATCTATTTTAGTGATAAAGCTTTAAGCTATATAAGCAGATACGA 1105
 Db 101 ValGlnAspValMetIle-----ArgLeuPro 109
 Qy 1106 GCCTCTCTCTAAATTAACGATCATGAGCTGGCGCTTATTGTGGGATTTGTCAGAAC 1165
 Db 110 AlaSerAsnAsnAspSerThrIle-----GlyGluHisValLysSerMetLeuGlnAsn 127
 Qy 1166 AGGCTAGATTTCTACGAAACTCTAAACGAAACGCAAAATTTTGGTCAAGGTAAAGC 1225
 Db 128 ValAspLysAspIle-----GlnIleArgSerIleGluPhe-----ValGly 141
 Qy 1266 AGCAAACTATCGAAGAAATCGTTTATCAGCGACCATCGGCTTTTAGGAGCTTTGGCA 1285
 Db 142 ProAsnValGlyGluGluLeuAlaGlnGlyAlaValTyrAlaThrLeuAlaThrLeuAla 161
 Qy 1286 ATCATCTTCTCTATGAGTTGGCTTTGAGTGGCAATATGCTTTCAGTGGCGCTATGC 1345
 Db 162 MetValLeuIleTyrValGlySerArgPheGluTrpArgLeuGlyPheGlySerIleAla 181
 Qy 1346 GCTTTAAATTCATGACCTTTTGCTACCTGTCAGCTGTTGTTTATAGCACATTTCTTTTG 1405
 Db 182 SerLeuAlaHisAspValIleThrLeuGlyVal-----PheSerAla 196
 Qy 1406 AAGAAATTCAAATAGATTGCAAGCCATTTGCTTTTAATGACTGTATTTGGGTTATTC 1465
 Db 197 LeuGlnIleGluIleAspLeuThrPheValAlaAlaIleLeuSerValValGlyTyrSer 216
 Qy 1466 TTAACAATACTTTCATCATTTTTCATCGTATTCGTGAAGAT---CGCAAGCGAACCTG 1522
 Db 217 IleAsnAspSerIleValValPheAspArgValArgGluAsnPheArgLysIleArgArg 236
 Qy 1523 TTTACCCCTATGCTGTTTATGATGCTTCAAAAGACCTTCAGCGCGACGTA 1582
 Db 237 LeuAspThrIleAspIle---IleAspIleSerLeuThrGlnThrLeuSerArgThrIle 255
 Qy 1593 ATGCAACAGCTACAACTATCATCTTTTGTAAATGCTTTTGTATAGGCGGCTCCTCT 1642
 Db 256 IleThrSerValThrThrLeuValValValMetAlaLeuPhePheGlyGlyProSer 275
 Qy 1643 GTCCTTAATTTTGCATTTATTATGACATAGGATTTCTTAGGAACTTTATCGTCTCTT 1702
 Db 276 IleHisAsnPheSerLeuAlaLeuLeuValGlyIleGlyPheGlyThrTyrSerSerIle 295
 Qy 1703 TATATTGCA 1711
 Db 296 PheValAla 298
 RESULT 21
 ID AAU91487 standard; Protein; 325 AA.
 XX
 AC AAU91487;
 XX
 DT 18-JUN-2002 (first entry)
 XX
 DE Haemophilus influenzae essential gene #65.
 XX
 KW Essential bacterial gene; antifungal agent; antibacterial agent;

KW antiparasitic agent; insecticidal agent; microbial infection;
 KW mucous membrane infection; otitis media; sinusitis; bronchitis;
 KW alveolitis; conjunctivitis; pneumonia; meningitis; epiglottitis;
 XX cellulitis; septic arthritis.
 OS Haemophilus influenzae.
 XX
 PN W0200218601-A2.
 XX
 PD 07-MAR-2002.
 XX
 PF 22-AUG-2001; 2001WO-US26245.
 XX
 PR 25-AUG-2000; 2000US-0649145.
 XX
 PA (ABBO) ABBOTT LAB.
 XX
 PI Chovan LE, Hessler PE, Reich KA;
 XX
 DR WPI; 2002-304258/34.
 DR N-PSDB; ABK64973.
 XX
 PT Essential bacterial genes in Haemophilus influenzae necessary for
 bacterium's growth and survival, useful for screening inhibitors of
 polypeptides and developing therapeutic agents e.g. antimicrobial
 PS Claim 9; Page 183-184; 185pp; English.
 XX
 CC The invention describes an essential bacterial gene (I) comprising a
 purified polynucleotide isolated from Haemophilus influenzae where (I)
 is essential to H. influenzae survival. The encoded polypeptide (II) is
 useful for screening substances that function to inhibit essential H.
 influenzae polypeptides by contacting (II) with the desired substances
 and measuring the response by a screen from specific, enzyme, general,
 affinity, phenotypic and binding screen. (I) and (II) are useful in
 developing therapeutic agents such as antifungal, antibacterial and
 antiparasitic agent, insecticidal agent, and preventive antimicrobial
 agents which are effective in preventing microbial infection or useful
 in treatment of that particular infection. (I) and (II) may also be
 useful in treatment of mucous membrane infections such as otitis media,
 sinusitis, bronchitis, alveolitis, conjunctivitis, pneumonia, meningitis,
 epiglottitis, cellulitis and septic arthritis. This is the amino acid
 CC sequence of an essential H. influenzae gene, described in the invention.
 XX
 SQ Sequence 325 AA;
 Alignment Scores:
 Pred. No.: 2,38e-29 Length: 325
 Score: 342.00 Matches: 98
 Percent Similarity: 51.90% Conservative: 80
 Best Local Similarity: 28.57% Mismatches: 111
 Query Match: 10.33% Indels: 54
 DB: Gaps: 15
 US-09-868-987-1 (1-1864) x AAU91487 (1-325)
 Qy 710 ATGACTAAATTTTCTTTCATGCTGCGATGAATAAGCAACATACACAGTTGCATATG 769
 Db 1 MetMetLysLeuPhe-----ThrLysAspLysAspGlyHisPhe 13
 Qy 770 ATGAATAAGTTCGTGGGATAAAG-----CATGATTTCTTGAGAGGATGC 814
 Db 14 IleArgGluIleAsnGlyIleLysLeuProPheProLeuThrGluPheMetLysValArg 33
 Qy 815 AAAAACTTTGGCGCTTTCTGGAGTGTTCCTTTTAGT-----TGGTGTGCT 865
 Db 34 LysLeuGlyTyrIleLeuSerAlaLeuLeuMetValIleSerLeuPheIleIleThr 53
 Qy 866 CTCGGGTTTGAGCCTCGAATTCGTTTGGGAATGGAATTTAAAGAGGATGATGCTTT 925
 Db 54 LysGlyPheAsn---Trp-----GlyLeuAspPheThrGlyGlyValPhe 68
 Qy 926 ACCTTTAATCCAAAGAGCATGGCATGCGATGTTGCTCAATCGTGGCAAGTTGTG 985

```

Db      69  AsptThrHisPheSerGlnSer-----AlaAsnLeuGlnGlnIleArgSer----- 83
Qy      986  CATPAACTACAGAAAGCTGCTCTTCTCTTCTAGAGACTTCCGTATCAACATTGGATCT 1045
Db      84  ---LysLeuHisGlnLeuAngLylIleGlnSerPro-----IleValGlnThrThrGlySer 100
Qy      1046  TCAGAAAAAGATCAAAATCTATTGATTAAGCTTAAGCTTACTAAGCAGATACGA 1105
Db      101  ValGlnAspValMetIle-----ArgLeuPro 109
Qy      1106  GCCCTCTCCATAAATTAAGCATGATGAGCTGCGCTTATTGGGATTTGGTACAGAAC 1165
Db      110  AlaSerAsnAsnAspSerThrIle-----GlyGlnHisValLysSerMetLeuGlnAsn 127
Qy      1166  AGGCTGATTTCTCTACGGAACCTTAACGAAACGCAAAATTTGGTCACAAAGTAAAC 1225
Db      128  ValAspLysAspIle-----GlnIleArgSerIleGluPhe-----ValGly 141
Qy      1226  AGCAAACTATCGAAAAAAGCGTTATCGCGCAACATCGGCGTTTAAAGAGCTTTGGCA 1285
Db      142  ProAsnValGlyGlnGlnLeuAlaGlnGlyAlaValTyrAlaThrLeuAlaThrLeuAla 161
Qy      1286  ATCATCTTCTCTATGATGAGCTTGGCTTGAATGGAATGCTTTCAGTCCGCTATGC 1345
Db      162  MetValLeuIleTyrValGlySerArgPheGlnTyrPargLeuGlyPheGlySerIleAla 181
Qy      1346  GCTTAATTCATGACCTTTTGGCTACCTGTCAGCTCTTGTATAGACATTTCTTTTGG 1405
Db      182  SerLeuAlaHisAspValIleIleThrLeuGlyVal-----PheSerAla 196
Qy      1406  AAGAAATTCATTAATGATTGTTCAAGCCATTGCTGTTAATGACTGTTTGGGATTTCA 1465
Db      197  LeuGlnIleGlnIleAspLeuThrPheValAlaAlaIleLeuSerValValGlyTyrSer 216
Qy      1466  TTAACATATCTTGATGATGATTTGGATGCTATGCTATGCTGAAGAT---CGCAACGCAACTG 1522
Db      217  IleAsnAspSerIleValValPheAspArgValAlaArgIleAsnPheArgLysIleArgArg 236
Qy      1523  TTTACCCCTATGATGATGTTTGTATATGATGCCCTTCAAAAGAGCTTACGCGCAGTA 1582
Db      237  LeuAspThrIleAspIle---IleAspIleSerLeuThrGlnThrLeuSerArgThrIle 255
Qy      1583  ATGACAAACAGCTACACTCTCATCAGTTTGTGTAATGCTTTTATAGCGGCTCTCT 1642
Db      256  IleThrSerValThrThrLeuValValValMetAlaLeuPhePheGlyGlyProSer 275
Qy      1643  GCTTTTATTTGATTTATATATGACCATAGGATTTCTTACGAACTTTTGCCTCTT 1702
Db      276  IleHisAsnPheSerLeuAlaLeuLeuValGlyIleGlyPheGlyThrTyrSerSerIle 295
Qy      1703  TATATTGCA 1711
Db      296  PheValAla 298

RESULT 22
AAU63631
ID AAU63631 standard; Protein; 576 AA.
XX
AC AAU63631;
XX
DT 27-FEB-2002 (first entry)
DE Propionibacterium acnes immunogenic protein #24527.
XX
KW SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
KW dermatological; osteopathic; neuroprotectant.
XX
OS Propionibacterium acnes.
XX
PN WC0200181581-A2.

```

```

XX      01-NOV-2001.
PD
XX
XX      20-APR-2001; 2001WO-US12865.
PF
XX      21-APR-2000; 2000US-199047P.
PR
XX      02-JUN-2000; 2000US-208841P.
PR
XX      07-JUL-2000; 2000US-216747P.
XX
XX      (CORI-) CORIXA CORP.
XX
XX      Skeiky YAM, Persing DH, Mitcham JL, Wang SS, Bhatia A;
PI      L'maisonneuve J, Zhang Y, Jen S, Carter D;
XX      WPI, 2001-616774/71.
DR      N-PSDB; AAS59634.
XX
XX      Propionibacterium acnes polypeptides and nucleic acids useful for
PT      vaccinating against and diagnosing infections, especially useful for
PT      treating acne vulgaris -
XX
XX      Claim 3; SEQ ID NO 24826; 1069bp; English.
PS
XX
XX      Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
CC      polypeptides. The proteins and their associated DNA sequences are used in
CC      the treatment, prevention and diagnosis of medical conditions caused by
CC      P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
CC      pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.
CC      P. acnes is also involved in infections of bone, joints and the central
CC      nervous system; however it is particularly involved in the inflammatory
CC      lesions associated with acne vulgaris. A method for detecting the
CC      presence or absence of P. acnes in a patient comprises contacting a
CC      sample with a binding agent that binds to the proteins of the invention
CC      and determining the amount of bound protein in the sample. The
CC      polypeptides may be used as antigens in the production of antibodies
CC      specific for P. acnes proteins. These antibodies can be used to
CC      downregulate expression and activity of P. acnes polypeptides and
CC      therefore treat P. acnes infections. The antibodies may also be used as
CC      diagnostic agents for determining P. acnes presence, for example, by
CC      enzyme linked immunosorbent assay (ELISA).
CC      Note: The sequence data for this patent did not form part of the printed
CC      specification, but was obtained in electronic format directly from WIPO
CC      at ftp.wipo.int/pub/published_pct_sequences.
XX
XX      SQ Sequence 576 AA;
XX
XX      Alignment Scores:
XX      Pred. No.: 3,45e-25 Length: 576
XX      Score: 306.50 Matches: 74
XX      Percent Similarity: 56.44% Conservative: 53
XX      Best Local Similarity: 32.89% Mismatches: 87
XX      Query Match: 9.25% Indels: 11
XX      DB: 22 Gaps: 3
XX
XX      US-09-868-987-1 (1-1864) x AAU63631 (1-576)
Qy      74  CGTATGCTGTAGTATGATGCGSTTATATGCTACGACGCCCTTATTTAAAC----- 124
Db      297  GlnPheAlaValValLeuAspGlyLysValIleSerSerProGlnLeuAngLysAsnThr 316
Qy      125  -----GTCCCATTTGAAA---AATCATGCCAGGTCTCCAGGAAATTTACCCACCGT 172
Db      317  GlyThrSerCysProIleAsnGlyGlyGlnAlaGlnIleSerGlyHisPheThrGlnAsn 336
Qy      173  GAAGTGAGCAAACTCGGCTCAGATTTTAAATCTGAGCGCATGCTTTTGTCCCGAGGTT 232
Db      337  SerAlaAlaAspLeuAlaAsnValLeuLysTyrGlyAlaLeuProLeuSerPheAspIle 356
Qy      233  CTCAGTGAAGACGATCTCTTCTGATCTTGGGAAAAACAATGATACCAAGCATATATC 292
Db      357  SerSerValAspAsnIleSerProThrLeuGlyGlyGlnGlnLeuArgAlaGlyIleIle 376
Qy      293  TCAGATGCTGCTGTGGCTTGCAATGCTTATTTGTTGATGAGCGGTATATTAATTTGGA 352

```


[illegible]

PI	Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
PI	Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX	WPI, 2001-376931/40.
DR	N-PSDB; AAH66795.
XX	
PT	Novel polynucleotides derived from Coryneform bacteria, for identifying
PT	mutation point of a gene, measuring expression of a gene, analysing
PT	expression profile or pattern of a gene and identifying homologous gene
XX	
PS	Claim 17; SEQ ID NO: 5330; 246pp + Sequence Listing; English.
XX	
CC	The present invention provides a number of nucleotide and protein
CC	sequences from the Coryneform bacterium Corynebacterium glutamicum. These
CC	are useful for identifying the mutation point of a gene derived from a
CC	mutant of coryneform bacterium, measuring expression amount and
CC	analysing the expression profile or expression pattern of a gene derived
CC	from Coryneform bacterium, and identifying a homologue of a gene derived
CC	from Coryneform bacterium. Coryneform bacteria are useful for producing
CC	amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC	particularly L-lysine. The present sequence is a protein described
CC	in the exemplification of the invention.
CC	Note: The sequence data for this patent did not form part of the printed
CC	specification, but was obtained in electronic format directly from the
CC	European Patent Office.
XX	
Q0	Sequence 644 AA;

Alignment Scores:	
Pred. No.:	7 13e-22
Score:	277.50
Percent Similarity:	51.61%
Best Local Similarity:	26.91%
Query Match:	8.38%
DB:	22
	Gaps: 7
	length: 644
	Matches: 337
	Conservative: 62
	Mismatches: 87
	Indels: 33

US-09-868-987-1 (1-1864) X AAC:91576 (1-644)

OY	5	ACGCCGATATTTCAGGAAGGGGATCAGCGGCACGTAAATGACAATAATTCGCAAC	64
Db	333	TTTTTTT :::-----	339
OY	65	CGTGATGCGCTATGCGCTGAGTGAATTCAGCGTAATATGTCAGACCCATTATTAAC	124
Db	340	-----GnlllealI..ethrleaApSerGlnValIlleSerlAlProvalIlegIn	356
OY	125	-----GTCCATTGAAAAMTCATGCCAGTGTCTCAGGAAATTATCCACCGTGAAGTG	178
Db	357	SerlAtnrPrOvalIglySerlAthrSerIlleThrGlyAspPheThrGlnThrGlnAla	376
OY	179	AGCAAACGCGCTCAGATTTAAMAATCGAGCG-----ATGCTTTTGTTCGCGAG--	229
Db	377	GlnaspLeuAlaasnIleauArgTyrglyAlaleuProleuSerPheIlaelyGluasn	396
OY	230	-----GTTTCAGTGAAGAAGCCGATCTTCTGATCTTGGGAAAAAACATGTACA	280
Db	397	GlyGluArgeIyglYthrTrfThrValProProSerleucIylalaIaSerleuys	416
OY	281	CMAGCATTATCTCAGCATCTGTCGCTTGGCAATGCTTATTCITTTGATGAGCTATAT	340
Db	417	IAlagIleuIlealIagIyleValIglyIleAlaleuValaIallePheValaPhealIayr	436
OY	341	TATGATTTGGAGGCGCATC-----GCTTCGGGAGACTGNTCTTCG	382
Db	437	TyrArgValPhegIlyPheValSerleuPheThrleuPheIlaIaIylaleuValItyr	456
OY	383	AATCTTTTGCTTATCTGGGACGCTCTACAGTATTTGGATCCGCCACTCACCTTGCAGA	442
Db	457	GlyleuIeuValleuLeuGly-----ArgTrpIleGlyTyserleuaspIeunIagly	474
OY	443	CTCGCTGGAGATTGTTCTTGCTATGCGGATGCGCGTAKAGTSCAAATTTCTTGATATCGAA	502
		::: TTTTTTT::: ::: ::: ::: ::: :::	

Dh 475 IlelaIdIyLeuIllelleglylleglyThrTrnAlaAapSerPheValValPheTyGlu 494

Oy 503 AGAATCCGAAGAAATTATTTATGTCTCAAAACTCTTAAAAATCTGTGAAAAACGATAT 562
|||||:::|||||:::|||||:::|||||:::
Db 495 ArgIleyAspGlnuleArggluGlYArGserPheaArgSerAlaValProArgAlaTrp 514
|||:::|||||:::|||||:::|||||:::
Oy 563 ACCAAGGCCTTTTGAGAGCACAATTTTGATTTCATTGACTACAGTAATGGCCCTCAGCAT 622
|||:::|||||:::|||||:::|||||:::
Db 515 GluSerAlaLysArgHrhlleValThrgLYamMetValThrLeuLeuGLYAlaIleVal 534
|||:::|||||:::|||||:::|||||:::
Oy 623 CTTTTCTCTCAATACAGGCGCTATTTAAAGGCTTGCTTGACACTG--ATTTTAGGA 679
:::|||||:::|||||:::|||||:::|||||:::
Db 535 IleYrXeuLeuAlaValgLyguVallysGlyPheAlaPheThrLeuGLYLeuThrThr 554
:::|||||:::|||||:::|||||:::|||||:::

Oy 680 ATTTTCTCTCAATGTTTACGGCTCT 706
:::|||||:::|||||:::|||||:::|||||:::

Db 555 ValPheAspLeuValValThrPheLeu 563
:::|||||:::|||||:::|||||:::|||||:::

RESULT 27
AAB05947 ID AAB05947 standard; Protein; 279 AA.
XX AAB05947;
XX
XX 20-OCT-2000 (first entry)
XX
DE Protein deduced from Mycoplasma hyopneumoniae genomic clone pAB727.
XX
XX Mycoplasma hyopneumoniae; vaccine; antigen; antimicrobial;
XX enzootic pneumonia.
XX
XX Mycoplasma hyopneumoniae.
XX OS WO200031115-A1.
PN
XX 02-JUN-2000.
XX
PF 19-NOV-1999; 99WC-AU01035.
XX
PR 20-NOV-1998; 98AU-0007273.
XX
PA (CSIR) COMMONWEALTH SCI & IND RES ORG.
XX
PI Moore RJ, Doran TJ;
XX WPI; 2000-400031/34.
DR N-P-SDB; AAA56833.
XX
PT Identifying antigenic polypeptides for use in vaccines against
PT Mycoplasma infection comprises screening proteins co-expressed with a
PT marker -
XX
PS Claim 27; Fig 14; 241pp; English.

The present sequence is encoded by a clone selected from a Mycoplasma hyopneumoniae genomic DNA library. M. hyopneumoniae causes Enzootic pneumonia in pigs. The infection rarely causes death but often results in significant depression, leading to reduced weight gain. The animals are also prone to secondary infection by opportunistic pathogens. Genomic DNA from M. hyopneumoniae was partially digested with restriction enzymes and fragments were ligated downstream of the polyHis tag region in the expression vector pCI. The ligation mixture was used to transform competent E. coli cells and polyHis positive clones were selected using anti-polyhis antibodies. Expression of the recombinant vector generates polyhis fusion proteins which are easily identified and isolated. Pigs were vaccinated with plasmid DNA from polyhis positive clones in order to identify clones encoding antigenic proteins which confer protection against M. hyopneumoniae infection. Such proteins may subsequently be used in the manufacture of effective vaccines.

Sequence 279 AA;

Alignment Scores:

Pred. No.: 1.94e-18 Length: 279
Score: 245.50 Matches: 77
Percent Similarity: 49.50% Conservatively: 73
Best Local Similarity: 25.41% Mismatches: 105
Query Match: 7.41% Indels: 49
DB: 21 Gaps: 9

US-09-868-987-1 (1-1864) x AAB05947 (1-279)

Qy 893 TTGGGAATGGATTTTAAAGAGGGGTATGCTTTTACCTTTAAATCCAAAGAGCATGGCATC 952
Dy 8 LeuAlaIleAAspPheSerGlyGlyThrAsnPheLeuIleGlu-SerSerAsnSerSery 27
Qy 953 AGCGATGTTGCTCAATCGTGGCGCAAGTGTGCATAACTACAGGAAGCTGCTTTCT 1012
Dy 27 rAspLeuIleThrLysGluLysAlaGluLysIleSerPhe----- 41
Qy 1013 TCTAGAGACTCCGTATTCAACATTTTCGATCTTCAGAAAGATCAAAATCTATTTAGT 1072
Dy 42 -----LeuAspSerGlnAsnIleAsnLysSerAsnSerTh 53
Qy 1073 GATAAGCTTTAAGCTATATAAGCAGATACGAGCTCTCTCTAAATTAACGATCATG 1132
Dy 53 rIleLeuLeu-----AsnProLeuAsnGluAsn----- 62
Qy 1133 AGCTGGCTTATTGTGGGATTGTTCTCAGAACAGCGCTAGATTCTCTACGGAA----- 1187
Dy 63 ---GlyAsnIlePheAsnLeuGluIleLysThrLysLeuAspLeuAlaThrLysIleAl 81
Qy 1188 -ACTCTAAACGNAACG-----CAAAATTTTGGTCA-----AAGGT 1221
Dy 81 aSerLeuAsnThrAlaIleGlnAsnAsnPheSerAsnIleArgMetThrAsnTyrSerIl 101
Qy 1222 AAGCAGCAAACTATCAAGAAATCGCTTATCAGCGACCATCGGGCTTTTAGGAGCTTT 1281
Dy 101 eSerAsnGluAlaGlnLysLeuIlePheAsnAlaIleLeuSerValGlyIleAlaLe 121
Qy 1282 GGCAATCATCTGCTCTATGTAGTTGCGCTTGAATGGCAATATGCTTTTCAGTGCCT 1341
Dy 121 uIlePheValThrIlePheThrLeuIleArgPheLysTrpThrPheSerLeuAlaIleIl 141
Qy 1342 ATGCGCTTAACTATGACCTTTGGCTACCTGTCAGTCTGTGTTATAGACATCTTCT 1401
Dy 141 ePheSerLeuLeuPheAsnValLeuMetValLeuLeuAlaIleIleThr----- 158
Qy 1402 TTTGAAGAAATTCAAATAGATTTCGAAGCATTCGTGCTTAATGACTGTATTGGCGTA 1461
Dy 159 -----ArgIleGluIleSerGlnAsnLeuValValAlaIleLeuThrIleGly 176
Qy 1462 TTCATTAACAATCTTTGATCATTTTGTATGCTATTCGT-----GAAGATCG 1509
Dy 176 rThrValAsnAspThrIleValValPheAspArgValLysAlaArgPheSerGluLeAs 196
Qy 1510 CCAACGGAACCTG-----TTTACCCCTATGATGCTTTTATGATGCTTCAAAA 1563
Dy 196 nHisGluAsnValTyrLysPheAspLysIleLysGluIleSerLeuGlnAlaIleArgGl 216
Qy 1564 GAGCTTCAGCGCAGGTAATGACACAGCTACACTCTATCAGTTTGTAAATGCTTTT 1623
Dy 216 uThrAlaLysArgSerValTyrThrSerLeuThrThrIleLeuThrIleValValLeuMe 236
Qy 1624 GTTTATAGCGGCTCTCTGCTCTTTAAATTTTGCATTTATATGACATAGGATCTTCT 1683
Dy 236 tIlePheTyrGluSerIleAspIleValPheSerLeuThrMetLeuIleGlyValIleIl 256
Qy 1684 AGGAACCTTATGCTCTTTATATGACCACCTCTGTGTGTGTATGTCGCGTAAGA 1743
Dy 256 eGlyThrTyrSerSerLeuPheIleAlaThrArgIleTrpIleLeuGluSerSerAr 276
Qy 1744 AAATCGC 1750
Dy 276 gAsnArg 278

RESULT 28
AAU63575
ID AAU63575 standard; Protein; 418 AA.
XX AC AAU63575;
XX 27-FEB-2002 (first entry)
XX DE Propionibacterium acnes immunogenic protein #24471.

XX KW SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
XX KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
XX KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
XX KW dermatological; osteopathic; neuroprotectant.

XX OS Propionibacterium acnes.

XX PN W0200181581-A2.

XX PD 01-NOV-2001.

XX PF 20-APR-2001; 2001WO-US12865.

XX PR 21-APR-2000; 2000US-199047P.

XX PR 02-JUN-2000; 2000US-208841P.

XX PR 07-JUL-2000; 2000US-216747P.

XX PA (CORI-) CORIXA CORP.

XX PI Skeiky YAM, Persing DH, Mitcham JL, Wang SS, Bhatia A;

XX PI L'Maisonneuve J, Zhang Y, Jen S, Carter D;

XX DR WPI; 2001-616774/71.

XX DR N-PSDB; AAS59634.

XX PT Propionibacterium acnes polypeptides and nucleic acids useful for

XX PS Claim 3; SEQ ID NO 24770; 1069pp; English.

XX CC Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic

XX CC polypeptides. The proteins and their associated DNA sequences are used in

XX CC the treatment, prevention and diagnosis of medical conditions caused by

XX CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,

XX CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.

XX CC P. acnes is also involved in infections of bone, joints and the central

XX CC nervous system, however it is particularly involved in the inflammatory

XX CC lesions associated with acne vulgaris. A method for detecting the

XX CC presence or absence of P. acnes in a patient comprises contacting a

XX CC sample with a binding agent that binds to the proteins of the invention

XX CC and determining the amount of bound protein in the sample. The

XX CC polypeptides may be used as antigens in the production of antibodies

XX CC specific for P. acnes proteins. These antibodies can be used to

XX CC downregulate expression and activity of P. acnes polypeptides and

XX CC therefore treat P. acnes infections. The antibodies may also be used as

XX CC diagnostic agents for determining P. acnes presence, for example, by

XX CC enzyme linked immunosorbent assay (ELISA).

XX CC Note: The sequence data for this patent did not form part of the printed

XX CC specification, but was obtained in electronic format directly from WIPO

XX CC at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 418 AA;

Alignment Scores:
Pred. No.: 4.85e-17 Length: 418
Score: 234.00 Matches: 77
Percent Similarity: 46.77% Conservatively: 75
Best Local Similarity: 23.69% Mismatches: 121
Query Match: 7.07% Indels: 52
DB: 22 Gaps: 13

US-09-868-987-1 (1-1864) x AAU53575 (1-418)

```

QY 788 ATAAAGCATGATTTCTTGAGAGAGATGCAAAAACCTTGGGCTGTTTCTGGAAGCTGTTT 847
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 66 lleserYrGlPhelleglAargArgysValTrpYrThrIleThralaValaIle 85
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 848 CTTTAGTTCGGCTGCTCGGGTTGGAGCCCTGGAATTCGGTTGGCAATGGATTT 907
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 86 lleserPheIleGlLeuValaIleValaIleValaIleValaIleValaIle 103
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 908 AAAGAGGGATGCTTACCTTAAATCAAGAGCATGCGATCGAGGATGCTCAA 967
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 104 LysGlyValaIleGlnPheArg---AlaProAlaIleValaIleThrSerThrIleAsp 122
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 968 ATGGGTGGCAAGTTGTCATAACTACAGAGAGCTGCTTTCTTCTAGAGAC--TTC 1024
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 123 ValArgLysSerValLeu-----SerSerGlyAlaProAspMetAspAlaThr 138
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1025 CGTATTCAACATTTGGATCTTCAGAAAAGATCAAAATCTTTTAGTGAATAGCTTTA 1084
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 139 GluValValSerLeuGlySerAsp-----AlaValGlnValGlnThrArgAlaLeu 155
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1085 AGC-----TATACATGACATAGATGAGCTCTCTCTCTAAATTAATGATCATGAGCTG 1138
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 156 AsnAsnAspGluThrThrLysValGlnGluAlaIleAlaLysAlaThr----- 171
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1139 CGTATTGTGGATGTGTGTCAGAAACAGGCTAGATTCTCTACGAACTCTAAACGA 1198
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 172 -----GlyThrLysMet 175
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1199 AAGCAAAATTTGGTCAAGCTAAGCAGCAACTATGAGAAAGAAATCCATTACAGCG 1258
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 176 SerSerValThrLysSerLysValGlySerGlnTrpGlyGlnIleThrGlyLysAla 195
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1259 ACCATGGGCTTTAGAGCTTTGGCAATCATCTGCTCTAGAGTTGCGCTT-- 1315
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 196 lIeLysAlaLeuValaIlePheLeuValLeuValMetLeuGlnIleTrpAlaLysPheArg 215
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1316 GAATGCAATATGCTTTCAGTCCGTATGCGCTTAAATCATGACCTTTGGTACCTGT 1375
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 216 HistTrpLysMetSerIleAlaLeuValaIleLeuLeuHisAspLeuIleValThrIle 235
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1376 GCACTCTGTTATAGCACTTTCTTTGGAAGAAATCAATGATGATTCAGCAACATT 1435
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 236 GlyIleLysAlaLeuValaIlePheThrValSerProSerThrVal-----Ile 251
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1366 GGTCTTAAATGATGCTTTCAGTCCGTATGCGCTTAAATCATGACCTTTGGTACCTGT 1495
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 252 GlyValLeu---ThrIleLeuGlyLysSerLeuLysTrpThrValaIlePheAspLys 270
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1496 ATTCGTGA-----GATGCCAAGGAACTGTTTACCCCTATGCAATGTT 1540
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 271 ValArgGluAsnValaIleAspIleAspLysArgAspLysThrPheAlaGluGly----- 288
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1541 TTAGTATATGATGCCCTTCAAAAGACGTTTCAGCCGACGAGTAAAGACAGCTACACT 1600
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 289 ---AlaAsnArgAlaValaIleAsnGlnValLeuValaIleSerIleAsnThrThrIleValaGly 307
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1601 CTATCAGTTTGTAAATGCTTTGTTT-----ATAGGGGCTCCCTCGTCTC 1645
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 308 ValLeuProValaIleAlaLeuLeuPheAlaGlyAlaPheValaIleGlySerGlyProLeu 327
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1646 TTTATTATTCATTATTAAGCATATGAGATTTCTTACGAACTTATGCTCTTAT 1705
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 328 GluAspLeuGlyLeuAlaLeuPheValaIleGlyMetIleValaIleLysSerSerIlePhe 347
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1706 ATTCGACCACTCTG 1720
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 348 lIleAlaThrProVal 352
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 29
AAG91574
ID AAG91574 standard; Protein; 403 AA.

```

XX AC AAG91574;
XX XX
XX 26-SEP-2001 (first entry)
XX DT
XX DE C glutamicum protein fragment SEQ ID NO: 5328.
XX XX
XX KW Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
XX KM organic acid synthesis.
XX OS
XX PN EPI108790-A2.
XX PD
XX XX
XX PF 20-JUN-2001.
XX PR 18-DEC-2000; 2000BP-0127688.
XX PR 16-DEC-1999; 993P-0377484.
XX PR 07-APR-2000; 2000UP-0159162.
XX PR 03-AUG-2000; 2000UP-0280988.
XX XX
XX PA (KYOW ) KYOWA HAKKO KOGYO KK.
XX PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
XX PI Tateishi N, Senon A, Ikeda M, Ozaki A;
XX DR N-PSDB; AAH6793.
XX DR WPI; 2001-376931/40.
XX PT Novel polynucleotides derived from Coryneform bacteria, for identifying
XX PT mutation point of a gene, measuring expression of a gene, analysing
XX PT expression profile or pattern of a gene and identifying homologous gene
XX PS
XX XX
XX Claim 17; SEQ ID NO: 5328; 246pp + Sequence Listing; English.
XX XX
XX The present invention provides a number of nucleotide and protein
XX sequences from the Coryneform bacterium Corynebacterium glutamicum. These
XX are useful for identifying the mutation point of a gene derived from a
XX mutant of coryneform bacterium, measuring expression amount and
XX analysing the expression profile or expression pattern of a gene derived
XX from Coryneform bacterium, and identifying a homologue of a gene derived
XX from coryneform bacterium. Coryneform bacteria are useful for producing
XX amino acids, nucleic acids, vitamins, saccharides and organic acids,
XX particularly L-lysine. The present sequence is a protein described
XX in the exemplification of the invention.
XX CC Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from the
XX European Patent Office.
XX CC
XX SQ
XX Sequence 403 AA;
XX XX
XX Alignment Scores:
XX Pred. No.: 2,286-16 Length: 403
XX Score: 228.00 Matches: 72
XX Percent Similarity: 44.97% Conservative: 80
XX Best Local Similarity: 21.30% Mismatches: 132
XX Query Match: 6.88% Indels: 54
XX DB: Gaps: 8
XX XX
XX US-09-868-987-1 (1-1864) x AAG91574 (1-403)
XX QY 797 GATTTCTTGAGAGATGCAAAAACCTTGGGCTGTTTCTGGAAGTGTCTTTTAGGT 856
XX ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 34 AspPheIleAlaLysThrLysLeuTrpLysTrpIleThrGlyLeuLeuValIleSer 53
XX QY 857 TGGCTTCCTCGGGTTTGGAGCCTGGAATTCGTTTGGGAATGATTTTAAAGAGG 916
XX ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 54 lIleuPheIleAlaIleArgLysPheSer-----LeuSerIleAspPheGlnGly 71
XX QY 917 TATGCTTACCTTATCAAAAGAGCATGAGCATGAGGATGTGCTAAATGCGTGGC 976
XX ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```


Db 72 ThrLysMetSerMetProAlaSerAspTyrSerThrGluGlnValGluGluThr----- 89
 Qy 977 AAAGTTGTCATAAATACTACAGAAAGCTGGTCTTCTTCTAGAGACTTCGCGATTCAACA 1036
 Db 90 -----PheThrGluAlaThrGlyIleThrProGluIleValGlnIleValGly 105
 Qy 1037 TTTCGATCTTCAGAAAGATCAAAATCTATTTTAGTGTAAAGCTTTAAGCTATACTAAG 1096
 Db 106 SerGlyAspAlaAargThrLeuGluIleTyr-----SerGluAargLeuSerAspGluAsp 123
 Qy 1097 CAGATACAGAGCT 1156
 Db 124 ValGluLysAlaAargLeu-----AlaIleTyr 132
 Qy 1157 GTCAGAAACAGGCTAGATTCTCTACGAAACTCTAAACGAAACGCAAAATTTTGGTCA 1216
 Db 133 GluGluTyrGlnProLeuAasnSerGluGlnProSerProAspAlaIleGlyAasnSer 152
 Qy 1217 AAGCTAAGCAGCAAACT 1276
 Db 153 ThrValSerGluSerTrpGlySerThrIleThrGlnArgMetValLeuAlaLeuIleAla 172
 Qy 1277 GCTTTGGCAATCATCT 1336
 Db 173 PheLeuValIleAlaAalleTyrIleAlaPheAargLeuGluAargGluMetAlaIleAla 192
 Qy 1337 GCGTATGCGCTTTAATTCATGACCTTTGGCTACTGTGTCAGTCTTGTATTATAGCACAT 1396
 Db 193 AlaMetAlaAlaLeuValValAsp-----GlyIleValIleAlaGlyIle 207
 Qy 1397 TCTTTTGGAGAAATCAATAGATTGCAAGCATTGGTCTTTAATGACCTGATTG 1456
 Db 208 TyrAlaValIleGlyLeuGluValSerProAlaThrValIleGlyLeuLeuThrValLeu 227
 Qy 1457 GGGTATTCATTAACAATPACTTGTGATCATTTTGTATCGTATTTCGTGAAGAT----- 1507
 Db 228 ThrPheSerIleTyrAspThrValValPheAspLysValArgGluAasnThrGluGly 247
 Qy 1508 -----CGCCAAGCGAACTGTTTACCCCTATGCAT 1537
 Db 248 PheGluGlySerArgArgThrTyrAlaGluGlnAlaAasnLeu----- 262
 Qy 1538 GTTTTAGTAATGATGCCCTTCAAAAGAGCTTCAGCCGACGGTAATGACAAACAGCTTACA 1597
 Db 263 -----AlaValAasnGlnThrPheMetArgSerIleSerThrIleIle 277
 Qy 1598 ACTCTATCAGTTTGTGTTAATGTTTGTATATA-----GGCGGCTCCTCT 1642
 Db 278 SerAlaLeuProIleIleAlaLeuMetValValAlaValTrpMetMetGlyValGlyThr 297
 Qy 1643 GTCTTTAATTTTGCATTTATTATGACCATAGGATTCCTTCTAGGAACCTTTATCGTCTCTT 1702
 Db 298 LeuLysAspLeuAlaLeuIleGlnLeuIleGlyValIleGluGlyThrPheSerSerVal 317
 Qy 1703 TATATTGACACACCTCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1756
 Db 318 PheLeuAlaThrProLeuLeuValSerLeuLysAsnArgLeuSerLysThrLys 335
 RESULT 30
 ID AAE15863 standard; Protein; 403 AA.
 AC AAE15863;
 DT 26-MAR-2002 (first entry)
 XX
 DE Corynebacterium glutamicum secF protein.
 KW Genetically modified bacterial strain; secD; secF; reporter system;
 OS enhanced secretion activity.
 OS Corynebacterium glutamicum.

PN WO200185967-A2.
 XX 15-NOV-2001.
 XX 26-APR-2001; 2001WO-EP04703.
 XX 12-MAY-2000; 2000EP-0110021.
 PA (DEGS) DEGUSSA AG.
 XX Berens S, Kalinowski J, Puehler A;
 DR WPI; 2002-082901/11.
 XX N-PSDB; AAD25591.
 DR
 XX Genetically modified Corynebacterium with enhanced secretion activity
 PT useful for production of desired substance e.g. protein, comprises a
 PT modification at one of the genes secD and secF -
 XX
 PS Claim 9; Page 41-42; 42pp; English.
 XX
 CC The present invention relates to genetically modified bacterial strain
 CC Corynebacterium glutamicum, comprising a genetical modification at one of
 CC the genes secD and secF. The genetically modified bacterial strain is
 CC useful for production of desired substance which is an amino acid,
 CC oligopeptide, polypeptide or protein preferably a heterologous protein,
 CC where the produced substance is secreted by the bacterial strain. The
 CC invention is useful in a reporter system. Modification in secD and secF
 CC in genetically modified bacterial strain Corynebacterium glutamicum,
 CC results in enhanced secretion of the strain, which is utilised for
 CC production of high amounts of desired substances which can be easily
 CC isolated from the source of production. The present sequence is
 CC Corynebacterium glutamicum secF protein.
 XX
 SQ Sequence 403 AA;
 Alignment Scores:
 Pred. No.: 2,28e-16 Length: 403
 Score: 228.00 Matches: 72
 Percent Similarity: 44.97% Conservative: 80
 Best Local Similarity: 21.30% Mismatches: 132
 Query Match: 6.88% Indels: 54
 DB: 23 Gaps: 8
 US-09-868-987-1 (1-1864) x AAE15863 (1-403)
 Qy 797 GATTCTTGAGAGATCAAAAACCTTTGGCTCTTCTCGAAGTGTCTTTCTTTTAGGT 856
 Db 34 AspPheIleAlaLysThrLysLeuTrpTyrTrpIleThrGlyIleLeuLeuValIleSer 53
 Qy 857 TGGCTTCTCTCGGGTTTGGAGGCTCGAAATTCGTTTGGGAATGGATTTTAAAGGAGGG 916
 Db 54 IleLeuPheIleAlaIleAargGlyPheSer-----LeuSerIleAspPheGlnGly 71
 Qy 917 TAGCCCTTTACCTTTAATCCAAAGACATGGCATCAGCATGTTCCTCAATCGGTGGC 976
 Db 72 ThrLysMetSerMetProAlaSerAspTyrSerThrGluGlnValGluGluThr----- 89
 Qy 977 AAAGTTGTCATAAATACTACAGAAAGCTGGTCTTCTTCTAGAGACTTCGCGATTCAACA 1036
 Db 90 -----PheThrGluAlaThrGlyIleThrProGluIleValGlnIleValGly 105
 Qy 1037 TTTCGATCTTCAGAAAGATCAAAATCTATTTTAGTGTAAAGCTTTAAGCTATACTAAG 1096
 Db 106 SerGlyAspAlaAargThrLeuGluIleTyr-----SerGluAargLeuSerAspGluAsp 123
 Qy 1097 CAGATACAGAGCT 1156
 Db 124 ValGluLysAlaAargLeu-----AlaIleTyr 132
 Qy 1157 GTCAGAAACAGGCTAGATTCTCTACGAAACTCTAAACGAAACGCAAAATTTTGGTCA 1216
 Db 133 GluGluTyrGlnProLeuAasnSerGluGlnProSerProAspAlaIleGlyAasnSer 152

QY 1217 AAGTAAGACAGCAACTATCAAGAAATGCGTTATCAGGACCATCGGCTTTAGGA 1276
 DB 153 ThrValSerGluSerTyrGlySerThrIleThrGlnArgMetValLeuAlaLeuIleAla 172
 QY 1277 GCTTTGGCAATCATCTCTCATGTGAGTTGGCTTGAATGGCAATATGCTTTCGT 1336
 DB 173 PheLeuValIleAlaIleIleTyrIleAlaPheArgLeuGlnArgGluMetAlaIleAla 192
 QY 1337 GCCGTATGCGCTTTAAATCATGATCCTTTGGCTACCTGTGACCTTGTATAGACAT 1396
 DB 193 AlaMetAlaIleAlaLeuValAlaAsp-----GlyIleValIleAlaGlyIle 207
 QY 1397 TTCCTTTGAAAGAAATTCATCAATAGATTTGCAAGCATTTGGCTTATATGCTATG 1456
 DB 208 TyrAlaValIleGlyLeuGlnValSerProAlaThrValIleGlyLeuLeuThrValLeu 227
 QY 1457 GGGTATTCATTAACAAATACCTTGATCATTGATCGTATTCGTATTCGTAAGAT----- 1507
 DB 228 ThrPheSerIleTyrAspThrValValAlaPheAspIysValArgGluAsnThrGluGly 247
 QY 1508 -----CGCCAAAGCAACCTGTTTACCCCTATGCAT 1537
 DB 248 PheGluGlySerArgArgGlyThrTyrAlaGluGlnAlaAsnLeu----- 262
 QY 1538 GTTTTAGTTAATGATGCCCTTCAAAAGCGTTACGCCGAGGTATATGACAAAGCTACA 1597
 DB 263 -----AlaValAsnGlnThrPheMetArgSerIleSerThrIleIle 277
 QY 1598 ACTGATCATGTTTGTATAGTCTTTGTTTATA-----GGCGGCTCCTCT 1642
 DB 278 SerAlaLeuProIleIleAlaLeuMetValAlaAlaValITPmetMetGlyValGlyThr 297
 QY 1643 GTCTTTAATTTTGCAATTTATATGACCATAGGATTCCTTACGAACTTATTCGTCTCT 1702
 DB 298 LeuIysAspLeuAlaLeuIleGlnLeuIleGlyValIleGluGlyThrPheSerSerVal 317
 QY 1703 TATATGCAACCACTCTGTGTGTGTTATGCTCCTAAAGAAATCGCTCAAAA 1756
 DB 318 PheLeuAlaThrProLeuAlaValSerLeuIysAsnArgLeuSerIysThrIys 335

RESULT 31

AAU36596 standard; Protein: 1055 AA.

AAU36596;

14-FEB-2002 (first entry)

Staphylococcus aureus cellular proliferation protein #766.

Antisense; prokaryotic cellular proliferation protein;

antibiotic; antibacterial; drug design.

Staphylococcus aureus.

WO200170955-A2.

27-SEP-2001.

21-MAR-2001; 2001WO-US09180.

21-MAR-2000; 2000US-191078P.

23-MAY-2000; 2000US-206848P.

26-MAY-2000; 2000US-207727P.

23-OCT-2000; 2000US-242578P.

27-NOV-2000; 2000US-253625P.

22-DEC-2000; 2000US-257931P.

16-FEB-2001; 2001US-269308P.

(ELIT-) ELITRA PHARM INC.

PI Yamamoto RT, Xu HH;
 XX WPI, 2001-611495/70.
 DR N-PSDB; AAS54455.
 XX New polynucleotides for the identification and development of
 PT antibiotics, comprise sequences of antisense nucleic acids -
 XX Example 3; Seq ID No 12189; 511pp, English.
 CC The invention relates to antisense inhibitors of genes essential to
 CC prokaryotic cellular proliferation, their use in identifying the
 CC genes, their use in the discovery of novel antibiotics, the essential
 CC genes themselves and the encoded proteins. The prokaryotes used are
 CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
 CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
 CC invention is also useful for the identification of potential new targets
 CC for antibiotic development. The antisense nucleic acids can also be used
 CC to identify proteins used in proliferation, to express these proteins,
 CC and to obtain antibodies capable of binding to the expressed proteins.
 CC The proteins can be used to screen compounds in rational drug discovery
 CC programmes. The antisense nucleic acid sequence is also useful to screen
 CC for homologous nucleic acids which are required for cell proliferation in
 CC a wide variety of organisms. The present sequence represents an
 CC essential prokaryotic cellular proliferation protein.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 1055 AA;
 Alignment Scores:
 Pred. No.: 5.18e-07 Length: 1055
 Score: 147.50 Matches: 132
 Percent Similarity: 34.42% Conservative: 111
 Best Local Similarity: 18.70% Mismatches: 229
 Query Match: 4.45% Indels: 234
 DB: 22 Gaps: 28
 US-09-868-987-1 (1-1864) x AAU36596 (1-1055)
 QY 265 GAAAGAAATGATGACCAAGCATTTATCTCAGATGCTGTGGCTGCAATGCTTATGT 324
 DB 366 GluIysSerLeuTyrThrMet-ValGluIysAlaSerLeuGlyThrIleValAlaIleI 385
 QY 325 TTGTATGAGCGGTATATTATAGA-----TTTGAGAGCGTCATCGCTTGGGACCTGTCT 378
 DB 385 eValIleLeuLeuPheLeuArgAsnIleArgThrThrAlaIleSerIleIleSerIlePr 405
 QY 379 TGTGAATCTTTTGGCTTATCTGCGCAGCTCTACAGTATTTGGATGGCGCACTCACCTTGT 438
 DB 405 OleSerLeuLeuMetAlaLeuIleAlaLeuIysLeuSerAspValSerLeuAsnIleIle 425
 QY 439 AGACCTGCTGGGATGTTGCTATGGGATGGCGGTATGCAATGCTTGTAT 498
 DB 425 uThrLeuGlyAlaLeuThrValAlaIleGlyArgValIleAspSerIleValValVa 445
 QY 499 CGAAGAAATC-----CGAGAGAAATTTTATGTCCTCAAGTCTTAA 540
 DB 445 IGIuAsnIleTyrArgArgLeuThrAspSerGluGlnGlnLeuIysGlyIuAsnLeu-- 464
 QY 541 AAAATCTGTGAAAAAGATATACCAAGGCTTTTGGAGCCATTTTGATCTTACACTTGAC 600
 DB 465 -----IleIleSerAlaThrThrGluValPheIysProIleMetSerSerThrIleVa 482
 QY 601 TACAGTATGCGCTCAGCATTTCTTTTCTTCAGATACAGGCGCTATT----- 649
 DB 482 IThrIleIleValPheLeuProLeuValPheVal---SerGlySerValGlyIuMetPh 501
 QY 650 -AAAGGTTGCTTGTGACATGATTTTATGAGATTTTCTTCAATGTTTACG----- 700
 DB 501 eArgProPheAlaLeuAlaIleAlaPheSerLeuLeuAlaSerLeuLeuValSerIleTh 521

PS Claim 7; Pages 1092-1095; 1657pp; French.

XX The present invention relates to the genomic sequence of *Pyrococcus*
 CC abysii (see AAF86431 and AAF41223-7) and *P. abysii* proteins. *P. abysii* is
 CC a hyperthermophilic archaeon, which is isolated from deep-sea
 CC hydrothermal vents. The present sequence is one such *P. abysii* protein.
 CC The proteins of the present invention have various potential industrial
 CC uses, since the proteins are stable at very high temperatures, some up to
 CC 110 degrees centigrade.
 CC Note: This patent is in the same patent family as WO200065062, which
 CC contains additional sequences as shown in AAB9132-AAB9143,
 CC AAF75903-AAH75920 and AAG66436.

CC Sequence 746 AA;

Alignment Scores:

Pred. No.:	4.9e-07	Length:	746
Score:	147.00	Matches:	119
Percent Similarity:	34.93%	Conservative:	115
Best Local Similarity:	17.76%	Mismatches:	258
Query Match:	4.44%	Indels:	178
DB:	22	Gaps:	24

US-09-868-987-1 (1-1864) x AAF66403 (1-746)

```

QY 23 GAGGGATCAGCGGCTGTAATGACATATTTGCAAAACCGTGAGTGGCTATGGCT 82
   |||
   |||
   |||
Db 138 GluAlaIArgAsnGlyLeuIleSerGlyAspTyr----- 148
QY 83 GTAGTATTCAGCGTTATATGTCAGAGCCCTATTTTAAACGTCCTCATG----- 133
   |||
   |||
   |||
Db 149 -----SerSerThrIleIleIleValIleAsnLeuAsnArgGlu 160
QY 134 AAAATGATGCGAGTGTCTCAGGAAATTAC-----CACGTGAGTGGAGCAATC 187
   |||
   |||
   |||
Db 161 LysAsnGlnLysAlaLeuValArgValTyrAsnAspIleGluArgGluIleGluArgThr 180
QY 188 -----GCCTAGATTAAATCTGAGCGAGTCTTTTGTCCGAGGTT 232
   |||
   |||
   |||
Db 181 AsnProGlnGlyValGluValValLeuThrGlyAspLeuGlyIleThrTyrLysIle 200
QY 233 CTCAGTGAAGAGAGATCTTTGATCTGTGAAATAAACAATGATACAGAGCATTC 292
   |||
   |||
   |||
Db 201 Leu-----GluMetLeuGlnAsnAspMetAsnArgThrMetAlaIleSerGlyIleIle 218
QY 293 TCAGATGCTGTGGCTGGCAATGCTTATTTGATGAGCGTATATTATAGA----- 346
   |||
   |||
   |||
Db 219 -----ValValLeuIleLeuLeuTyrPheTyrLysSerPro 230
QY 347 -----TTTGAGGCGTCATCGCTTCGGAGCT 373
   |||
   |||
   |||
Db 231 IleArgMetLeuValProLeuValProLeuIlePheGlyValIleMetThrLeuGlyPhe 250
QY 374 GTTCTTCGATCTTTTGTCTTATCTGGCAGCTCTACAGTATTGGATGGCCCACTACC 433
   |||
   |||
   |||
Db 251 MetGlyLeu-----LeuGlyIleProLeuAsp 259
QY 434 TTTCAGAGCTCGTGGAGATTGTTCTTGCTATGGGAGTGGCGGTAGATGCAATGTTCTT 493
   |||
   |||
   |||
Db 260 IleAlaThrThrThrValGlyAlaMetIleIleGlyMetGlyIleAspTyrGlyValHis 279
QY 494 GTATTTCGAAGAATCCGAGAGATTTTATGTTCTCAAGTCTTAAATAATCTGTAGAA 553
   |||
   |||
   |||
Db 280 ValThrAsnArgTyrTyrGluGluArgGlyLysGlyLysSerLeuGluAlaIleGlu 299
QY 554 AAAGATATACAGAGCTTTTGAGCATTTTGTATCTAATCTGATACAGATTTGGCC 613
   |||
   |||
   |||
Db 300 GluAlaIleAlaGluThrGlyLysAlaLeuLeuGlyAlaAlaIleThrThrIleAlaGly 319
QY 614 TCAGACACTCTTTCTTCTAGATACAGAGGCTATTTAAAGGTTTCTTGGATTTGATTT 673
   |||
   |||
   |||
Db 320 PheLeuAlaLeuSerIleSerIleLeuProSerIleuYsaGluSerValSerIleVal 339

```

```

QY 674 TTAGATTTTCTTCATATGTTTACGGCTTTTTCATGACTAAATTTTCTTCATGCTG 733
   |||
   |||
   |||
Db 340 MetGlyLeuGlyLeuAlaIleAsnAlaValIleValThrProAlaLeuAlaIleLeu 359
QY 734 TGGATGATTAAGACCCATACATACAGTTCATGATGATGATTAATGTTCTGGGATTAAG 793
   |||
   |||
   |||
Db 360 -----GluGluGluPheArgGlyLysValMetLys----- 370
QY 794 CATGATTTCTGAGAGATGCAAAAACTTGCGCTGTTCTCGAAGT----- 841
   |||
   |||
   |||
Db 371 -----LysGluIleIleAlaIleGlyGlyLysSerLysIle 383
QY 842 -----GTTTCTTTAGTTCGCTGCTCGGAGTTTGAGCCGTAATTCGTTTG 895
   |||
   |||
   |||
Db 384 SerPheIlePheSerMetLeuGlyLysAlaIleLysArgSerProThrPheAlaLeuLeu 403
QY 896 -----GGAATGATTTTAAAGAGGATGATGCTTTATCTTTAAT--- 934
   |||
   |||
   |||
Db 404 IleAlaLeuIleIleSerGlyValSerLeuTyrGlyAlaSerLysIleThrThrGluVal 423
QY 935 -----CCAAAGACATGCGATCGCATGCTTGTCT 964
   |||
   |||
   |||
Db 424 ArgLeuGluLysMetIleProThrAspLeuProGluIleGluAlaLeuSerAspIleArg 443
QY 965 CAATGCGTGGC----- 976
   |||
   |||
   |||
Db 444 SerGluPheGlyGlyGluAspGluValThrIleLeuIleLysAlaAspValArgAsp 463
QY 977 -----AAAGTTGTG 985
   |||
   |||
   |||
Db 464 ProThrLeuValArgAspIleLeuArgPheGluArgGluIleLysAlaAspSerTyrIle 483
QY 986 CATTAACATACAGAGAGCTGCTTTCTTCTAGAGACTTCGATTTAAACATTTGGA--- 1042
   |||
   |||
   |||
Db 484 AsnAsnValPheGluThr---GlnSerIleAlaAspValValIleGlnLysTyrGlyTyr 502
QY 1043 -----TCTTCAGAAAAGATC-----AAATCTATTATTAAGT 1072
   |||
   |||
   |||
Db 503 IleProGluAspLysGlyLysIleSerGluAlaIleGluGlySerSerLeuValSerSer 522
QY 1073 GATTAAGCTTTAAGCTATCTAATGACAGATACAGAGCTCTCTCTAAATTAACG----- 1126
   |||
   |||
   |||
Db 523 AspTyrSerMetThrIleIleLys---LeuLysGlyAsnPheMetGlyValThrGlnSer 541
QY 1127 -----ATCATGAGCTGGCGCTTATGCGGATTTGCGATTTGTCGAAAGAGCGCTAGA 1174
   |||
   |||
   |||
Db 542 GluPheAsnArgIleMetGluTyrPheGluArgAlaIle-----GlnArgAlaAsp 558
QY 1175 TTT-----CTTACGGAACCTTAACGAACCAAAATTTTGG 1213
   |||
   |||
   |||
Db 559 PheProGluValLysValGluLeuAlaGlyLysSerTyrLeuAsnTyrValIleAsn 578
QY 1214 TCAAAGTAAAGCAAACTATGAAAGAAATGCGTTATACAGCGACATCGGGCTTTTA 1273
   |||
   |||
   |||
Db 579 AspLeuValAsnGluGluLeuGlyLys-----IleSerThrIle 591
QY 1274 GAGCTTTGGCAATCATCTTGCTCTATGAGAGTTTGGCTTTGAATGCAATATGCTTC 1333
   |||
   |||
   |||
Db 592 GlyThrLeuIleValAlaMetValIlePheAlaIlePheArgArgProThrValSerIle 611
QY 1334 AGTGGCGTATGCGCTTAAATTCATGACCTTTTGGTACCTGTCACATCTTGTATTAGCA 1393
   |||
   |||
   |||
Db 612 AlaMetIleMetProMetPheLeuGlyAlaIleThrThr-----IleGly 626
QY 1394 CATTTCTTTTGAAGAAATTCAAATAGATTGCAAGCAATGTCCTTTAATGACTGTA 1453
   |||
   |||
   |||
Db 627 TyrMetGlyLeuAlaGlyIleProPheSerGlnThrLeuAlaGlyValAlaSerMetIle 646
QY 1454 TTGGGATTAATCAATTAACATATCTTGATGATTTTGGATGTATTTGGTGAAGATCCGCA 1513
   |||
   |||
   |||
Db 647 ValGlyLeuGlyValAlaAspTyrGlyMetHisIleThrHisArgPheLeuGluGluAsn 666
QY 1514 GCGAAGCTGTTTACCCCTATGATGTTTATGATGATGATGATGATGATGATGATGATG 1573

```

```

Db 667 GluGlyAsnLysThrPro-----:::|||||:::
Qy 1574 CGCAGGTAATGACACAGCTACACTCTATCAGTTTGTAAATGCTTTTCTTTATAGGC 1633
Db 663 GlyIlePheValGlyAlaLeuThrThrAlaGlyGlyPheLeuAlaLeuSerAlaGln 702
Qy 1634 GGCTCCTCTGCTTTAAATTTGCAATTATTATGACCATAGGATTCCTCTAGGAACTTTA 1693
Db 703 LeuThrAlaIleHisAspPheGlyArgValLeuAlaValGlyIlePheAlaSerMetPhe 722
Qy 1694 TCGTCTCTTATATGACACCACTCTGTTG 1723
Db 723 AlaAlaTyrLeuValThrProAlaIleLeu 732

RESULT 33
ID AAY35183 standard; Protein; 564 AA.
XX
AC AAY35183;
XX
DT 13-SEP-1999 (first entry)
XX
DE C. pneumoniae protein involved in metabolism of nucleic acids.
XX
KW Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;
KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;
KW vaccine; neutralising epitope.
XX
OS Chlamydia pneumoniae.
XX
PN WO9927105-A2.
XX
PD 03-JUN-1999.
XX
PF 20-NOV-1998; 98WO-IB01890.
XX
PR 04-NOV-1998; 98US-0107078.
PR 21-NOV-1997; 97FR-0014673.
XX
PA (G8ST ) GENSET.
XX
PI Griffiths R;
XX
DR WPI; 1999-357842/30.
XX
PT Genome sequence of Chlamydia pneumoniae
XX
PS Page 1041-1042; Disclosure; 1912pp; English.
XX
CC AAY34584-Y35879 represent the proteins encoded by all the open reading
CC frames in the complete genome (see AAX91990) of Chlamydia pneumoniae.
CC C. pneumoniae causes respiratory disease such as pneumonia and
CC bronchitis and is thought to be a contributing factor in heart
CC disease, sarcoidosis, sinusitis, purulent otitis media, erythema
CC nodosum or pharyngitis. The polypeptides encoded by the open reading
CC frames of the C. pneumoniae genome (see AAY34584-Y35879) can be used in
CC immunogenic compositions as vaccines. Vectors containing C. pneumoniae
CC nucleotide sequences can also be used as immunogenic compositions,
CC especially where the vector directs the expression of a neutralising
CC epitope of C. pneumoniae.
XX
SQ Sequence 564 AA;

Alignment Scores:
Pred. No.: 7,1e-07 Length: 564
Score: 145.00 Matches: 25
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.38% Indels: 0
DB: 20 Gaps: 0

US-09-868-987-1 (1-1864) x AAY35183 (1-564)

```

```

Qy 1788 CAATATATAAATCTCCTTTGGGACTTTAGTCCCAAGGCCCTGCTGTATATAATTATG 1847
Db 1 GlnTyrIysAsnLeuLeuTrpAspPheSerProIysGlyProCybGlyIleLysPheMet 20
Qy 1848 ACAAAATTCACATAAT 1862
Db 21 ThrAsnSerAspAsn 25

RESULT 34
AAU34738
ID AAU34738 standard; Protein; 1034 AA.
XX
AC AAU34738;
XX
DT 14-FEB-2002 (first entry)
XX
DE E. coli cellular proliferation protein #319.
XX
KW Antisense; prokaryotic cellular proliferation protein;
KW antibiotic; antibacterial; drug design.
XX
OS Escherichia coli.
XX
PN WO200170955-A2.
XX
PD 27-SEP-2001.
XX
PF 21-MAR-2001; 2001WO-US09180.
XX
PR 21-MAR-2000; 2000US-191078P.
PR 23-MAY-2000; 2000US-206848P.
PR 26-MAY-2000; 2000US-207727P.
PR 23-OCT-2000; 2000US-242578P.
PR 27-NOV-2000; 2000US-253625P.
PR 22-DEC-2000; 2000US-257931P.
PR 16-FEB-2001; 2001US-289308P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Haselbeck R, Ohlssen KL, Zyekind JW, Wall D, Trawick JD, Carr GJ;
PI Yamamoto RT, Xu HH;
XX
DR WPI; 2001-611495/70.
DR N-PSDB; AAS52597.
XX
PT New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids -
XX
PS Example 3; Seq ID No 10331; 51pp; English.
XX
CC The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the
CC genes, their use in the discovery of novel antibiotics, the essential
CC genes themselves and the encoded proteins. The prokaryotes used are
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence represents an
CC essential prokaryotic cellular proliferation protein.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1034 AA;

```

Alignment Scores:

Pred. No.:	1,28e-06	Length:	1034
Score:	144.00	Matches:	135
Percent Similarity:	31.73%	Conservative:	89
Best Local Similarity:	19.12%	Mismatches:	226
Query Match:	4.35%	Indels:	256
DB:	22	Gaps:	30

US-09-868-987-1 (1-1864) x AAU34738 (1-1034)

```

QY 308 TTGGCAATGCTTATGTTTATGATGCGTATATATAGATTGGAGGCGTCGCTTCG 367
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 353 LeuValMetYrLeuPheLeuGlnAsnMetArGlaThrLeuIleProThrIleAlaVal 372
QY 368 GGAGCTGTTCTTGAATCTTTGCTTATCGGCGAGCTCTACAGTATTTGGAGCGCA 427
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 373 ProValIleLeuGlyThrPheAlaIleLeuAlaIlePheGlyTyrSerIleAsnThr 392
QY 428 CTCACCTTGTACAGACTCGCGGATTTGCTGTATGGAGTGGCGGTAGTCAAAAT 487
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 393 LeuThrMetPheGlyMet-----ValLeuAlaIleGlyLeuLeuValAspAspAla 409
QY 488 GTTCCTGTATTCGAAGAATCCGAGAGGAATTTTATGTCTCAAGTCTT-----AAA 541
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 410 IleValIleValIleGluAsnVal---GluArgValMetMetGluAspLysLeuProDlys 428
QY 542 AAATCTGTAGAAAAGATATACCAAGGCTTTGGAGCCATTTTGAATTTGAATCTGACT 601
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 429 GluAlaThrGluLysSerMetSerGlnIleGlnIleAlaLeuVal-----GlyIleAla 446
QY 602 ACACTATTTGGCTCAGCACTTCTT-----TTCTCTTCAGATATACAGGCTTAT 649
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 447 MetValIleSerAlaValAlaPheIleProMetAlaPhePheGlyGlySerThrGlyAlaIle 466
QY 650 ---AAAGGTTGCTTGTACATTTGATTTAGAAATTTTCTCTCAATGTTTACGCTCTT 706
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 467 TyrArgGlnPheSerIleThrIleValSerAlaMetAlaLeuSerValLeuValAlaLeu 486
QY 707 TTCATGACT----- 715
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 487 IleLeuThrProAlaLeuCyAlaThrLeuLeuLysProValSerAlaGluHisIleGlu 506
QY 716 ---AAATTTTCTTCATCCCGTGAATGATTAAGACC---CAACATACACAGTTGCATATG 769
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 507 AsnLysGlyGlyPhePheGlyTyrPheAsnThrThrPheAspHisSerValAsnHisTyr 526
QY 770 ATGAAT-----AAGTCGTGGGATTAAGCATGATTTCTTG----- 805
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 527 ThrAsnSerValIleGlyIleLeuGlySerThrGlyArgIleLeuLeuIleTyrAlaLeu 546
QY 806 -----ACAGATGCAGAAAACCTTGGGCTGTTCTGGAAGTGTCTTCTT----- 850
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 547 IleValIleGlyMetValIleLeuPheLeuArgLeuProSerSerPheLeuProGluGlu 566
QY 851 -----TTAGTTCGCTGCTCGGCTTTGAGGCTGGAAT----- 886
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 567 AspGlnGlyValPheLeuThrMetIleGlnLeuProAlaGlyAlaThrGlnGluArgThr 586
QY 886 ----- 886
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 587 GlnLysValIleAspGlnValIleThrAspTyrTyrLeuLysAsnGluLysAlaAsnValGlu 606
QY 887 -----TCCGTTTGGGAATGATTTTAAAGA----- 913
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 607 SerValPheThrValAsnGlyPheSerPheSerGlyGlnIleGlnAsnAlaGlyMetAla 626
QY 913 ----- 913
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 627 PheValSerLeuLysProTyrGlnGluArgAsnGlyAspGluAsnSerAlaGluAlaVal 646
QY 914 -----GGGTATGCTTATACCTTAT 934
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 647 IleHisArgIleLysMetGluLeuGlyLysIleArgAspGlyPheValIleProPheAsn 666

```

```

QY 934 ----- 934
Db 667 MetProAlaIleValIleLeuGlyThrAlaThrGlyPheAspPheGluLeuIleAspGln 686
QY 934 ----- 934
Db 687 AlaGlyLeuGlyHisAspAlaLeuThrGlnAlaArgAsnGlnLeuLeuGlyMetAlaAla 706
QY 935 -----CCAAAAGACATGCATCGACGATGCTGCTCAAAATG 970
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 707 GlnHisProAlaSerLeuValSerValArgProAsnGlyLeuGluAspThrAlaGlnPhe 726
QY 971 CGTGGCAAGTT---GTGCATTAACCTACAGGAAGCTGGTCTTCTTCTAGACCTCCGT 1027
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 727 LysLeuGlnValAspGlnGluLysGlnAlaLeuGlyValSerLeuSerAspIleAsn 746
QY 1028 ATTCAACATTTGATCTTCA----- 1048
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 747 ---GlnThrIleSerThrAlaLeuGlyGlyThrTyrValAlaAsnAspPheIleAspArgGly 765
QY 1049 -----GAAAAGATCAAAATCTATTTAGTATATAA 1078
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 766 ArgValLysLeuTyrValGlnAlaAspAlaLysPheArgMetLeuProGluAspVal 785
QY 1079 GCTTTAAGCTATACCTACAGCAGATACGAGCTCTCTCTTA-----AAATTAAGATC 1129
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 786 AspLysLeuTyrValArgSerAlaAsnGlyGluMetValProPheSerAlaPheThrThr 805
QY 1130 ATGAGCTG-----CGTTATTTGGGATTT---GTTCACGA 1162
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 806 SerHisThrValTyrGlySerProArgLeuGluArgTyrAsnGlyLeuProSerMetGlu 825
QY 1163 AACAGGCTTAGATTTCTTACGAAACTTAACGAAGCA---AAATTTGCTCAAG 1219
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 826 IleGlnGlyGluAlaIleProGlyThrSerSerGlyAspAlaMetAlaLeuMetGluAsn 845
QY 1220 GTAAAGCAGCAACTATTCG-----AAGAAATGCGTTTCAAGCG 1258
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 846 LeuAlaSerLysLeuProAlaGlyIleGlyTyrAspTyrThrGlyMetSerTyrGlnGlu 865
QY 1259 ACCATCGCG-----CTTTAGAGACTTTGGCAATCATCTGCTCATATGTG 1303
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 866 ArgLeuSerGlyLysGlnAlaProAlaLeuValAlaIleSerPheValValAlaPheLeu 885
QY 1304 AGTTGGCTTTGAATGCAATATGCTTTCAGTCCGATAGCCTTAAATCATGACCTT 1363
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 886 CysLeuAlaIleLeuTyrGlnSerTyrSerIleProValSerValMetLeuValAlaPro 905
QY 1364 TTGGCTAACCTGTCAGCTCTTGTATATACACATTTCTTTTGAAGAAATTCAAATAGAT 1423
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 906 LeuGlyIleValIleValIleLeuLeuAlaIleThrLeuPheAsnGlnLys-----AsnAsp 923
QY 1424 TTCCAAGCCATGTGCTTTAATAGCTGATTTGGGCTATTCATTAAACAATACCTTGATC 1483
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 924 ValTyrPheMetValGlyLeuLeuThrThrIleGlyLeuSerLysAlaSerAlaIleLeu 943
QY 1484 ATTTTGATCGTATTCGTGAAGATCCCAAGCGACACTGTTTACCCCTATGATGTTTAA 1543
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 944 IleValGlu-----Phe 947
QY 1544 GTTAATGATGCCCTTCAAAAG-----ACGTTACAGCCGACGGTAATGACACA 1591
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 948 AlaLysAspLeuMetGlnLysGlyLysGlyValValGlnAlaThrLeuMetAlaVal 967
QY 1592 GCTAACACTATACAGTTTGTATATG-----CTTTGTTTATA----- 1630
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 968 ArgMetArgLeuArgProIleLeuMetThrSerLeuAlaPheIleLeuGlyValLeuPro 987
QY 1631 -----GGCGGCTCTCTGCTTAAATTTTGCATTTATATGACCATTA 1672
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 988 LeuAlaIleSerAsnGlnAlaGlySerGlyAlaGlnAsn-----AlaValGlyIle 1004

```

QY 1673 GGGATTCCTTAGACCTTATCGTCTCTTTATATGCA-----CCACCTCTGTG 1723
 Db |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
 QY 1005 GlyValMetGlyMetValSerAlaThrLeuLeuAlaIlePheValProValPhe 1024
 QY 1724 TTGTTTATGGTCCGTA 1741
 Db |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
 QY 1025 PheValValIleArgArg 1030
 Db |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
 RESULT 35
 AAG82041
 ID AAG82041 standard; Protein; 178 AA.
 AC AAG82041;
 XX
 DT 03-SEP-2001 (first entry)
 DE S. epidermidis open reading frame protein sequence SEQ ID NO:1176.
 KW Staphylococcus epidermidis SR1 strain; infection; diagnosis;
 KW vaccination; endocarditis.
 OS Staphylococcus epidermidis.
 XX WO200134809-A2.
 XX 17-MAY-2001.
 XX 09-NOV-2000; 2000WO-US30782.
 XX 09-NOV-1999; 99US-0164258.
 XX (GLAX) GLAXO GROUP LTD.
 XX Kimmerly WJ;
 XX WPI; 2001-316495/33.
 XX N-FSDB; AAH52891.
 XX Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
 XX useful for vaccinating against infections, e.g. endocarditis -
 XX Claim 18; Page 337; 2189pp; English.
 XX AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides
 XX (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis.
 XX (I) and (II) can have antibacterial activity and therefore can be used
 XX in vaccination. The nucleic acids (I) may be used to produce the
 XX S. epidermidis polypeptides (II) via the production of vectors
 XX containing them which are used to produce hosts cells which express the
 XX polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
 XX used to vaccinate subjects and to raise antibodies against the bacteria.
 XX The polypeptides may also be used to assay for other inhibitors of their
 XX activity and therefore identify compounds that may be used for the
 XX treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to
 XX AAH55030 represent specifically claimed S. epidermidis genomic DNA
 XX polynucleotide sequences from the present invention. AAH55091 to
 XX AAH55098 represent oligonucleotide sequences and primers which are used
 XX in the exemplification of the present invention.
 XX N.B. The present invention specifically claims all the polynucleotide
 XX sequences given in the sequence listing of the present specification,
 XX however the sequence listing only goes up to SEQ ID NO:4454 so even
 XX though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,
 XX no sequences are present for SEQ ID NO:4455 to 4464.
 XX Sequence 178 AA;
 Alignment Scores:
 Pred. No.: 1.41e-06 Length: 178
 Score: 140.00 Matches: 47
 Percent Similarity: 46.41% Conservative: 24
 Best Local Similarity: 30.72% Mismatches: 58
 Query Match: 4.38% Indels: 24
 DB: 22 Gaps: 4

US-09-868-987-1 (1-1864) x AAG82041 (1-178)
 QY 801 AAATCATGCTTTATCCCGACCACTTATTCATCATCATGCAACTGTGTATGTTGGTCTTA 742
 Db |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
 QY 32 LysSerCys-----ThrSerLeuProSerLeuLeuSerCysLeuSerPhePheLeu 48
 Db |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
 QY 741 -----TTCATC-----CAC 733
 Db |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
 QY 49 ThrProAsnAsnGlnTyrCysPheLeuLeuLysPheGluAspThrSerAsnAspAsnAsn 68
 Db |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
 QY 732 AGCATGAAGAAATTTAGTCATGAAAAGAGCCCGTAAACATTTGAAGAGAAATTCCTAAA 673
 Db |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
 QY 69 ProLeuAspLysAsnThrAlaValThrAsnIlele-----LysIleProLys 84
 Db |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
 QY 672 ATCAATGTAAGCAACCCCTTTAATAGGCCCTGTATCTAGGAAGAAAAGAGTGCAG 613
 Db |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
 QY 85 SerAsnMetValAlaLysProLeuThrGluLeuSerProLysLysLysSerThrAlaAla 104
 Db |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
 QY 612 GCCAATAGTGTAGTCAAGTTAGAAATCAAAATGGCTCCAAAAGCCTTGGTATATCCTTTT 553
 Db |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
 QY 105 AlaMetThrValValLysLeuGluSerAsnIleValLysAsnGluLeuLeuPheAlaPhe 124
 Db |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
 QY 552 TCTACAGATTTTAAAGACTTTGAGACAAATAAAATTCCTCTCGGATTCCTCGAATACA 493
 Db |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
 QY 125 GluTyrAlaCysLeuSerValArgProIleLeuSerSerSerLeuIleArgSerTyrIle 144
 Db |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
 QY 492 AGAATTTTGCATCTACGGCCATCCCATAGCAAGAAACAATCCCGAGCTCTGACAAAG 433
 Db |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
 QY 145 MetIleLeuAlaSerThrAlaIleProThrProLysThrAsnAlaAlaAsnProGlyArg 164
 Db |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
 QY 432 GTGAGTGGCGCATCCAAATACTGTAGAGCTGCCAGATA 394
 Db |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
 QY 165 ValAsnThrProAspMetLysLeuAsnAlaThrLysVal 177
 Db |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
 RESULT 36
 ABB49954
 ID ABB49954 standard; Protein; 744 AA.
 AC ABB49954;
 XX
 DT 05-FEB-2002 (first entry)
 DE Listeria monocytogenes protein #2658.
 KW Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;
 KW vitamin B12; bacterial infection; disease.
 XX Listeria monocytogenes.
 OS
 XX WO200177335-A2.
 XX 18-OCT-2001.
 XX 11-APR-2001; 2001WO-FR01118.
 XX 11-APR-2000; 2000FR-0004629.
 XX (INSP) INST PASTEUR.
 XX Buchrieser C, Frangeul L, Couve E, Rusniok C, Faihi H, Dehoux P;
 XX Dussurget O, Chetoui F, Nedjari H, Glaser P, Kunst F, Cossart P;
 XX Daniels J, Goebel W, Kref J, Kuhn M, Ng E, Vazquez-Boland JA;
 XX Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A;
 XX Chakraborty T, Domann E, Hain T, Berche P, Charbit A, Durant L;
 XX Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N;
 XX Maduenio E, De Pablo B, Wehland J, Kaerst U, Entian K, Hauf J;
 XX Rose M, Voss H;
 XX WPI; 2002-010914/01.
 XX Genomic sequence for Listeria monocytogenes, useful e.g. for treatment
 XX and prevention of Listeria and related bacterial infections, and
 PT

PT related polypeptides -
 XX Claim 6; SEQ ID No 2659; 192bp; French.
 PS
 XX
 CC The present invention relates to the genome sequence of *Listeria*
 CC monocytogenes EGD-e (see ABA03041). The genome sequence and fragments of
 CC it are useful for selecting probes and primers for detecting genes in *L.*
 CC monocytogenes and related organisms, and for studying genetic
 CC polymorphisms and other genomes. The present sequence is a protein
 CC encoded by the genome sequence of the present invention. Proteins
 CC expressed from the genome sequence are useful for raising specific
 CC antibodies, identification of *L. monocytogenes* and related organisms, and
 CC for biosynthesis and biodegradation, especially biosynthesis of Vitamin
 CC B12. The genome sequence and proteins encoded by it are also useful for
 CC selecting compounds that regulate gene expression and cell replication
 CC and modulate *L. monocytogenes*-related diseases. In addition, the genome
 CC sequence and proteins encoded by it are useful in pharmaceutical and
 CC vaccines compositions for the treatment or prevention of infections by *L.*
 CC monocytogenes and related organisms.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 744 AA;
 Alignment Scores:
 Pred. No.: 3.04e-06 Length: 744
 Score: 140.00 Matches: 126
 Percent Similarity: 35.80% Conservative: 106
 Best Local Similarity: 19.44% Mismatches: 232
 Query Match: 4.23% Indels: 184
 DB: Gaps: 29
 US-09-868-987-1 (1-1864) x ABE49954 (1-744)
 QY 125 GTCCATGAAATCATGCGAGTCTCAGG----- 157
 Db 145 ValAlaLeuLysGlnGlnValSerLysAspGlyThrPheIleGlnThrValIleMet 164
 QY 158 -----AAATTACCAACCTGGAAGTAGAGCAAACTCCCTCAGATTAAATCGAGCG 211
 Db 165 LysAspSerIleThrSerAspGlnLeuThrGlnSerPheLysGlnLeuGlnSerAlaAla 184
 QY 212 ATGCTTTTGTCCGAGGTTCTCAGTAGAGACGATCTTTTGAT-----CTT 262
 Db 185 LysThrThrIleGlyGlnAspProPheAlaGlnValSerAlaAspAspThrLeuVal 204
 QY 263 GGGAAAAACATGTACACAAAGGAT----- 289
 Db 205 AlaArgThrThrGlyProAlaGlyIleSerValAspAlaSerGlyLeuPheLysAspAla 224
 QY 290 ---ATTCAGCATGCTGTGCTTGCAATGCTTAT---GTTTGTAGAGGTTATATTAT 343
 Db 225 AspValSerLeuLeuIleGlyThrValLeuLeuValPheLeuLeuValIleTyr 244
 QY 344 AGATTGAGGCGTCATGCTTCGAGAGCTGTTCTTCTGAATCTTTTGCTTATCTGGCA 403
 Db 245 Arg---SerProIleLeuAla-----LeuIleProLeuIleAlaValGlyPhe 259
 QY 404 GCTTACAGTATTGGATGCGCCACTGACCTGTACAGACTCGCTGGGATT----- 454
 Db 260 AlaIleLeuValIleThrProIleGlyLeuLeuAlaLysGlnGlyIleIleThrTyr 279
 QY 455 -----GTTCTGTCTATGGGATGCGCTGATGATGCAAT 487
 Db 280 GlySerGlnGlyLeuSerIleMetThrValLeuLeuPheGlyAlaGlyThrAspTyrCys 299
 QY 488 GTTCTGTATTGAAAGATCCGAGAGGAATTTTATTGTTCTCAAGCTTAAATAATCT 547
 Db 300 LeuPheLeuIleSerArgPheArgSerHisLeuHisThrGlnLysAspAlaGlyPheGlnAla 319
 QY 548 GTAAAAAAGATATACCAAGCGTTTGGAGCCATTTTGTGATTCACTTACATACAGTA 607
 Db 548 -----ValPheIleValAlaLeuGlyGlnAsp-----TyrAsnIlePheMet 649

Db 320 PheLysGlnAlaPheSerGlyThrAlaGlyAlaIleAlaLeuSerGlyLeuThrValMet 339
 QY 608 TTGGCTCAGCACTTTTCTTCCAGTATACAGGCGCTATTAAGGCTTGCTTGACA 667
 Db 340 AlaAlaLeuLeuLeuLeuAlaAlaGlnTyrGlySerPheHisAspPheAlaValPro 359
 QY 668 TTGATTGTGAATTTTCTTCAATGTTTACGGCTTTTCATGACTAAATTTTCTTC 727
 Db 360 PheSerLeuAlaIlePheIleMetCelleSerSerLeuThrLeuValProAlaLeuLeu 379
 QY 728 -----ATGCTGTGG----- 736
 Db 380 GlyIlePheGlyArgValSerPheThrProPheValProArgThrValGlnMetGlnGlu 399
 QY 737 -----ATGAATTAAGCCCACTACACAGTTGCATATGATATAGTTC----- 781
 Db 400 ThrArgAlaLysLysLysGlySerLysThrProLysHisSerLysGlnAsnArgPheThrHis 419
 QY 782 -----GTGGGATTAAGCATGATTTCTTGAGAGATGCAAAACCTTGGGCT 829
 Db 420 LysIleGlyGlnMetSerAlaLysHisProVal-----ArgIleLeuIle 434
 QY 830 GTTCTGGAAGTGTCTTTTCTTTAGGTTGCTGTCTCGGTTTGAGCCTGGAAT--- 886
 Db 435 IleThr---LeuIleIleLeuIleGlyCys-----GlyIlePheThrThr 448
 QY 887 -----TCCGTTTGGAGATGATTTTAAAGAGGCTAT 919
 Db 449 GlnValLysTyrThrTyrAspThrLeuSerThrPheProGlnAspMetProSerArgGlu 468
 QY 920 GCCTTACCTTAAATCAACAAAGCATGAGCATCAGGATGTTGCTCAAAATGCGTGCAAA 979
 Db 469 GlyPheThrLeuIleSerAspHisPheGlyAlaGlyMetLeuAlaProMet-----Glu 486
 QY 980 GTTGTGCATTAACCTACAGAGAGCTGCTCTTCTTACAGAC-----TTCGATTTCAA 1033
 Db 487 ValValValAsnSerLysGlnSerMetLysSerSerLeuGlnAsnValAsnGlyValAla 506
 QY 1034 ACATTTGATCTTCAGAAAAGATCAAAATCTAT----- 1066
 Db 507 SerValThrGlyProGlnArgSerLysGlyTyrGlnLysTyrThrLeuIleLeuLysAsp 526
 QY 1067 -----TTTAGATTAAGCTTTTAACTATACCTAATACAGATTCAGGCTCTCTCAAAA 1120
 Db 527 AsnProTyrSerAsnGlnAlaMetAspValValProLysValArgAlaAla----- 543
 QY 1121 TTACGATCATGAGCTGGCGTTATTGTGGATGTTGTCAAAACAGGCTGATTTCTC 1180
 Db 544 -----AlaAspLysGlyAsnAspVal 550
 QY 1181 TACGAAACTTAAACGAAACGAAATTTTGTCAAAAGTAAAGCAAACTATCGAAG 1240
 Db 551 TyrIleAlaGlyGlnThrAlaThrGlnTyrAspAspArgAlaValThrGlnHisAspGlu 570
 QY 1241 AAAATGCTTATCAGCGCAACATGGGCTTTTAAAGGCTTTGCAATCATCTGCTCAT 1300
 Db 571 LysVal-----IleIleProLeuValIleIleLeuIleAlaIleLeuLeu 586
 QY 1301 GTAGTTTGGCTTGAATGGAATATGCTTTCAGTCCGCTATGCGCTTAAATTCATGAC 1360
 Db 587 CysTyrLeuArg-----SerIleThrAlaMetLeuTyr--- 597
 QY 1361 CTTTGGCTTACCTGTGCAGCTTGTATTATACCATTTCTTTTGAAGAAATTCAAATA 1420
 Db 598 LeuValAlaThrValLeuLeuSerPheValGlyAlaLeuGlyLeuGlyTyrValIleIle 617
 QY 1421 -----GATTGACGCACTTGGTGTCTTAAATGCTATGAGGATTCATTA 1468
 Db 618 HisTyrAlaMetGlyValGlnAlaIleSerGlyLeuIleProLeuTyrAlaPhe----- 635
 QY 1469 AACAAATCTTGTATCATTTTGTATCGTATTCGTAAGATCGCCAAAGCACTGTT--- 1525
 Db 636 -----ValPheIleValAlaLeuGlyGlnAsp-----TyrAsnIlePheMet 649

Qy 1526 -----ACCCCTATGCGATGTTTATGTTAATGATGCC 1555
 Db 650 IleSerSerIleTrpLysAsnSerLysLysMetProLeuArgLysAlaIleThrGluGly 669
 Qy 1556 CTTCAAAAGACGTTCCAGCGCAGGTAATGACACAGCT----- 1594
 Db 670 ValGlyGlnThrGlyGly-----ValIleThrSerAlaGlyLeuIleLeuAlaGlyThr 687
 Qy 1595 -----ACAACCTATCATGTTTGTAAATGCTTTTGTATAGCGGCTCCTCT 1642
 Db 688 PheGlyValLeuThrThrLeuProIleGlnLeuVal----- 700
 Qy 1643 GTCCTTAATTTGCAATTTATATGACCATAGGATCTTCTAGGAACATTTATCGTCTCT 1702
 Db 701 -----GlnPheGlyLeuIleThrAlaIleGlyValLeuLeuAspThr----- 714
 Qy 1703 TATATTGACCACTCTCTGTTG 1726
 Db 715 PheIleValArgPropheLeuVal 722
 RESULT 37
 AAB96433
 ID AAB96433 standard; Protein; 507 AA.
 XX
 AC AAB96433;
 XX
 DT 29-OCT-2001 (first entry)
 XX
 DE Putative P. abyssi preprotein translocase subunit #3.
 XX
 KW Hyperthermophilic archaeon; hyperthermophilic protein.
 XX
 OS Pyrococcus abyssi.
 XX
 PN FR2792651-A1.
 XX
 PD 27-OCT-2000.
 XX
 PF 21-APR-1999; 99FR-0005034.
 XX
 PR 21-APR-1999; 99FR-0005034.
 XX
 PA (CNRS) CNRS CENT NAT RECH SCI.
 XX
 PI (IFRE)- IFREMER INST FR RECH EXPL MER.
 XX
 PI Forterre P, Thierry JC, Prieur D, Dietrich J, Lecompte O;
 PI Querellou J, Weissenbach J, Saurin W, Heilig R;
 XX
 DR WPI; 2001-126236/14.
 XX
 XX New nucleotide sequences isolated from Pyrococcus abyssi encode
 PT proteins useful in industry -
 XX
 FS Claim 7; Pages 1128-1129; 1657pp; French.
 XX
 CC The present invention relates to the genomic sequence of Pyrococcus
 CC abyssi (see AAF86431 and AAH41223-7) and P. abyssi proteins. P. abyssi is
 CC a hyperthermophilic archaeon, which is isolated from deep-sea
 CC hydrothermal vents. The present sequence is one such P. abyssi protein.
 CC The proteins of the present invention have various potential industrial
 CC uses, since the proteins are stable at very high temperatures, some up to
 CC 110 degrees centigrade.
 CC Note: This patent is in the same patent family as WO2000065062, which
 CC contains additional sequences as shown in AAB99132-AAB99143,
 CC AAH75903-AAH75920 and AAG66436.
 XX
 SQ Sequence 507 AA;
 Alignment Scores:
 Pred. No.: 2.82e-06 Length: 507
 Score: 139.50 Matches: 47
 Percent Similarity: 51.46% Conservative: 41

Best Local Similarity: 27.49% Mismatches: 70
 Query Match: 4.21% Indels: 13
 DB: 22 Gaps: 5
 US-09-868-987-1 (1-1864) x AAB96433 (1-507)
 Qy 197 TTAATACTGGAGGAGTCTTTGTTCCGAGGTTCTCAGTGAAGACAGATCTCTTCT 256
 Db 318 LeuArgSerGlySerLeuProValLysLeuSerIleGluArgIleAspTyrIleSerPro 337
 Qy 257 GATCTTCGGAAAAACAATGTACACAAGGATTTATCTCAGCATCTGTGGCTTCGCAATG 316
 Db 338 LysLeuGlyGluAspPheLysArgGlnValIleAlaGlyIleAlaLeuLeuVal 357
 Qy 317 CTATTCTTTTGTAGGAGCTATATTATAGA-----TTTGGAGGCGTCATC 361
 Db 358 ValGlyAlaIleValTyrLeuHisTyrArgLysLeuLysIleAlaIleProValMetPhe 377
 Qy 362 GCTTCGGAGCTGTTCTTGAATCTTTTGTCTTATCTGGGAGCTCTACAGTAT---TTG 418
 Db 378 ThrSerPheSerGluValLeuIleLeuGlyIleAlaSerIleIleArgTrpAsnLeu 397
 Qy 419 GATCGGCACCTCACCTGTGTGAGGACTCGCTGGGATTTCTTGTATGGGATGCGCGTA 478
 Db 398 AspLeuPro-----SerIleAlaGlyIleIleAlaIleGlyThrGlyVal 413
 Qy 479 GATCAAAATGTTCTTGTATTCGAAAGAAATCCGAGAGGAATTTTATTGCTCAAAGT--- 535
 Db 414 AspGlnGlnIleValIleThrAspGluLeuLeuGlyAspValThrAlaGlyLysLysArg 433
 Qy 536 ---CTTAAAAAATCT---GTAGAAAAAGGATATACCAAGGCTTTTGGAGCCATTTTGAT 589
 Db 434 IleValLysArgSerGlyIleLeuLysArgMetGlyArgAlaPhePheIleLeuAla 453
 Qy 590 TCTAACTTGATCAGATATTGGCTCAGCAGCTTCTTTCTTCTTAGATACAGGCGCTATT 649
 Db 454 SerAlaSerThrThrIleValAlaMetSerPheLeuPheLysPhePheValGlyGlyLeu 473
 Qy 650 AAAGGGTTGCTTTGACATTTGATTTTAGGAATT 682
 Db 474 ArgGlyPheAlaPheThrThrIleLeuGlyIle 484
 RESULT 38
 AAB96433
 ID AAB96433 standard; Protein; 1098 AA.
 XX
 AC AAB96433;
 XX
 DT 18-FEB-2002 (first entry)
 XX
 DE Novel human diagnostic protein #30358.
 XX
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 XX
 OS Homo sapiens.
 XX
 XX WO200175067-A2.
 PN
 XX
 PD 11-OCT-2001.
 XX
 PF 30-MAR-2001; 2001WO-US08631.
 XX
 PR 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 XX Drmanac RT, Liu C, Tang YT;
 XX
 DR WPI; 2001-639362/73.
 DR N-PSDB; AAS94554.
 XX


```

Db 744 AlaMetTyrLeuCysGlyPheSerLeuAsnLeu-----SerLeuMet 758
Qy 1439 GCTTTAATGACTGTTATGGGTATTCATTAAACAATACTTTTGATCATTTTTCATGCTGATT 1498
Db 759 AlaLeuThrIleAlaThrGlyPheValValAspAlaIleValValLeuGluAsnIle 778
Qy 1499 CGTAGATCCCAAGCAACCTGTTTACCCTATGCTATGCTATGTTTGTATGATGATGCTT 1558
Db 779 AlaArgHisLeuGluAlaGlyMet---LysProLeuGlnAlaLeuGlnGlyThrArg 797
Qy 1559 CAAAGAGCTTTCAGCCGACGCTAATGACACAGCTACACTCTA---TCAGTTTGTGTA 1615
Db 798 GluValGlyPhe-----ThrValLeuSerMetSerLeuSerLeuValAlaValPheLeu 815
Qy 1616 ATGCTTTTGTATTATAGGCG-----GGCTCCTCTGCTTAAATTTTGCAATTTATTATG 1666
Db 816 ProLeuLeuLeuMetGlyGlyLeuProGlyArgLeuLeuArgGluPheAlaValThrLeu 835
Qy 1667 ACCATAGGATCTCTTAGGAACCTTATCGCTCTTATATATGTCACACCTCTGTTGTG 1726
Db 836 SerValAlaIleGlyIleSerLeuLeuValSerLeuThrLeuThrProMetCysGly 855
Qy 1727 TTTATGGTCCGT 1738
Db 856 TrpMetLeuLys 859

RESULT 39
AAV19957
ID AAV19957 standard; Protein; 1007 AA.
XX
AC AAV19957;
XX
DT 19-JUL-1999 (first entry)
XX
DE B. burgdorferi antigenic protein, t814.aa.
XX
KW Antigenic protein; vaccine; Lyme disease; infection; detection.
XX
OS Borrelia burgdorferi.
XX
PN W05859071-A1.
XX
PD 30-DEC-1998.
XX
PF 18-JUN-1998; 98WO-US12718.
XX
PR 03-SEP-1997; 97US-0057483.
XX
PR 20-JUN-1997; 97US-0050359.
XX
PR 22-JUL-1997; 97US-0053344.
XX
PR 22-JUL-1997; 97US-0053377.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PA (MEDI-) MEDIMUNE INC.
XX
PI Choi GH, Erwin AL, Hanson MS, Lathigra R;
XX
WPI; 1999-189980/16.
XX
DR N-PSDB; AAX61654.
XX
PT New isolated Borrelia burgdorferi nucleic acids - used to develop
PT products for the diagnosis, prevention and treatment of diseases
PT caused by Borrelia, particularly Lyme disease
XX
PS Claim 12; Page 135; 275pp; English.
XX
CC This sequence represents a Borrelia burgdorferi (Bb) protein of the
CC invention, which is suitable for use in a vaccine. The Bb polypeptides
CC can be used in vaccines for eliciting protective antibodies to members of
CC the Borrelia genus, particularly for the use against Lyme disease in
CC humans and animals. They can be used for preventing or attenuating an
CC infection caused by a member of the Borrelia genus. The products can also
CC be used for detection of members of the Borrelia genus.

```

```

XX SQ Sequence 1007 AA;
Alignment Scores:      8.93e-06      Length:      1007
Pred. No.:      136.50      Matches:      128
Score:      32.97%      Conservative:      117
Percent Similarity:      17.23%      Mismatches:      243
Best Local Similarity:      4.12%      Indels:      255
Query Match:      20      Gaps:      29
DB:
US-09-868-987-1 (1-1864) x AAV19957 (1-1007)
Qy 173 GAAGTAGCAAACTC-----GCCTCAGATTTAAATCTGGAGCGATCTCTTT 220
Db 281 GluIleGluLysLeuLysLeuSerMetProLysAspMetLysLeu----- 295
Qy 221 GTTCCCAGAGTCTCTCAGTGAAGAGACGATCTCTCTCATCTTGGGAAAAAACATGTACA 280
Db 296 -----GluIleAlaSerAsp-----SerThrAspPheIleLysAlaSerIleSer 310
Qy 281 CAAGGCATTATCTCAGCATGCTGTGGCTTGGCAATGCTTATTGTTTTCATCAGCGTATAT 340
Db 311 ThrValValAsnSerAlaTyrPheGlyAlaMetLeuAlaIlePheValIlePhePhePhe 330
Qy 341 TATAGA---TTTGAGGCGTCATCGCTTCGGGA---GCTGTCTCTCTGAATCTTTTCTT 394
Db 331 LeuArgSerPheArgAlaThrIleIleIleGlyIleSerIleProIleAlaIleValLeu 350
Qy 395 ATCTGGCAGCTCTACAGTATTGGATGCCCACTACCTTGTCCAGGACTCGCTGGGATT 454
Db 351 ThrPheCysLeuMetTyrPheValAsnIleSerLeuAsnIleMetSerLeuAlaGlyLeu 370
Qy 455 GTTCTTCTATGGGCGCGTAGATGCAAAATGTTCTGTATTTCGAAAGAAATCCGAGAG 514
Db 371 AlaLeuGlyIleGlyMetValValAspCysSerIleValIleAspAsnIleTyrLys 390
Qy 515 GAATTTTATTGTCTCAAGCTTTAAATAATCTGTAGAAAAAGATATACCAAGCGTTT 574
Db 391 TyrArgGlnLysGlyAlaLysLeuIleSerSerIleLeuGlyAlaGlnGluMetMet 410
Qy 575 GGAGCCATTTTGTATTCTTAACCTTGACTACA-----GTATTGGCTCAGCACTTCTTTC 628
Db 411 LeuProIleThrSerSerThrPheThrSerIleCysValPheGlyProPheLeuIlePhe 430
Qy 629 TTCCTAGATACAGGCGCT-----ATTAAGGGTTTGTCTTTGACATTCATTATA 676
Db 431 LysSerGluLeuGlyValTyrGlyAspPhePheLysAspPheThrPheThrIleValIle 450
Qy 677 GGAATTTTCTTCAATGTTTACGGCTCTTTTCATGCTAAATTTTCTTCATCTCTGG 736
Db 451 SerLeuGlyValSerLeuLeuValAlaIlePheLeuValProValLeuSerSerHisTyr 470
Qy 737 ATGATTAAGACCCCAACATACACAGTTCATATGATGATTAAGTTTCGTGGGG----- 787
Db 471 ValGlyLeuTyrThrSerPheGlnLysAsnIleLysAsnAlaPheIleArgLysIleAsp 490
Qy 787 ----- 787
Db 491 AlaPhePheAlaSerIleTyrTyrPheLeuGluPheLeuTyrIleAsnLeuLeuAsnIle 510
Qy 788 ---ATAAGCATGATTTCTTTGAGAGGATGCAAAAAACTTTTGGGCTGTTTCTGGAAGTGT 844
Db 511 ValLeuAsnHisLysLeuIlePheGlyLeuIleValPhePheSerPheIleGlySerLeu 530
Qy 845 -----TTTCTTTTAGGTTGGCTCTCGGGTTTGGAGCGCTCGGAATTCGTT--- 892
Db 531 LeuLeuGlyLeuLeuLeuAspValThrThrPheThrArgGlyLysGluAsnSerIleThr 550
Qy 893 TTGGGAATGATTTT----- 907
Db 551 IleAsnLeuAsnPheProHisLysThrAsnLeuGluTyrAlaLysPheTyrSerAsnArg 570

```

QY 908 -----AAAGAGCGTAT----- 919
 Db 571 PheLeuGluIleValIysSerGluAlaIysGlyTyrIysSerIleIleAlaThrLeuArg 590
 QY 920 -----GCCTTACCTTAAATCA-----AAAGACATGCGATCAGATGTT 961
 Db 591 AlaAspArgIleThrPheAsnValLeuPheProLeuIysGluIleuSerArgAspAsnLeu 610
 QY 962 GCTCA-----ATGCGTGCAAGTGTGCATTAACCTACAGAAAGCT 1003
 Db 611 ThrGlnSerValAspTyrAspSerIleIysTyrIysIleMetAsnArgIle----- 627
 QY 1004 GGTCTTCTTACAGACTTCCGATTCAA-----ACATTGGATCT 1045
 Db 628 GlyAsnLeuTyrProGluPheAsnIleGluProSerIleSerGlyAsnAlaLeuGlyGly 647
 QY 1046 TCAGAAAAGATCAAAATCATATTAGTATTAAGCTTAAGCTTACTAAG----- 1096
 Db 648 GlyAspSerIleIysIleIysIleSerAlaAsnAspPheGluTyrIleIysAspTyrGly 667
 QY 1097 CAGATACGAGCCTCTCTCTAAATTAACGATC----- 1129
 Db 668 LysIleLeuValSerMetLeuIysGluIleProGluLeuValAsnProArgLeuSer 687
 QY 1130 -----ATGAGCTGGCGCTTAT 1144
 Db 688 IleSerAspPheGlnLeuGlnIleGlyValGluIleAspArgAlaLeuValTyrAsnTyr 707
 QY 1145 -----TGCGATTTGTTGTC 1159
 Db 708 GlyIleAspMetAsnThrIleLeuAsnGluLeuIysAlaAsnIleAsnGlyValAla 727
 QY 1159 ----- 1159
 Db 728 GlyGlnTyrValGluIysGlyLeuAsnTyrAspIleValLeuIysLeuAspArgMetAsp 747
 QY 1159 ----- 1159
 Db 748 ValIysAsnLeuIysAspLeuGluIysIlePheIleThrAsnSerSerGlyValIysIle 767
 QY 1160 -----AGA 1162
 Db 768 ProPheSerSerIleAlaThrPheGluIysThrAsnIysAlaGluSerIleTyrArgGlu 787
 QY 1163 AACAGCCTAGATTCTCTCGAAGAAC-----CTTAA 1195
 Db 788 AsnGlnAlaLeuThrIleTyrLeuAsnIleGlyIleSerProAspAspAsnLeuThrGln 807
 QY 1196 CGAAACGCAAAATTTGGTCAAGGTAAAGCAAGCAACTATGCAAGAA----- 1243
 Db 808 ValThrAlaIysValIAspPheIleAsnAsnIysValProHisIysGluGlyIleThr 827
 QY 1244 -----ATGCGTTATCAGCGCAGCATCGGG 1267
 Db 828 LeuIysValGluGlyGluTyrAsnGluPheSerAsnIleMetAsnGlnPheIleIle 847
 QY 1268 CTTTAGAGCCTTGGCAATCATCTTGCTATGAGTGTGCGTTGAATGGCAATAT 1327
 Db 848 IleMetMetAlaIleIleValIAspPheGlyIleMetAlaSerGlnPheGlu----- 864
 QY 1328 GCTTACGTGCGCGTATTAATCATGACCTTTGGCTACTGTGCGACTGTGTT 1387
 Db 865 SerPheLeuIysProPheIleIleIlePheThrIleProLeuThrAlaIleGlyValVal 884
 QY 1388 ATACACATTTCTTTTGG-----AGAAATTCOAATAGATTGCAAGCCATGGCTTTA 1444
 Db 885 LeuIleHisPheLeuAlaGlyGluIysLeuSerIle-----PheAlaAlaIleIleIleu 903
 QY 1445 ATGACTGATTGGGATTATCATTAACAATCTTGATCATTTTGTATTCGTGAA 1504
 Db 904 MetLeuVal-----GlyValValValAsnThrGlyIleValLeuValAsp----- 918
 QY 1505 GATGCGCAAGCAACTGTTTACCCCTATGCATGTT----- 1540

Db 919 -----TyrThrGlyLeuLeuIleIysArgGlyPheGlyLeuArgGlu 932
 QY 1541 TTAGTAAATGATGCTCCCTTCAAAAGACGTTACGGCGATTAATGACACAGCTACAAT 1600
 Db 933 AlaIleIleGluSerCysArgSerArgLeuArgProIleuMetSerSerLeuThrSer 952
 QY 1601 CTATCAGTTTGTATATGCTTTGTTTATAGGGCGCTCCCT-----GCTTTAAT 1651
 Db 953 IleIleGlyLeuIleProMetAlaPheSerSerGlySerGlyAsnGluLeuIysPro 972
 QY 1652 TTTGCAATTTATTAATGACATAGGATTTCTCTAGAACTTATGCTCTTATATTGCA 1711
 Db 973 IleAlaPheThrPheIleGlyGlyMetThrAlaSerThrPheLeuThrIlePheIle 992
 QY 1712 CCACCTGTG 1720
 Db 993 ProMetLeu 995
 RESULT 40
 AAU5552
 ID AAU5552 standard; Protein; 1032 AA.
 XX AAU5552;
 AC
 XX 14-FEB-2002 (first entry)
 DT
 XX
 XX Haemophilus influenzae cellular proliferation protein #193.
 DE
 XX Antisense; prokaryotic cellular proliferation protein;
 XX antibiotic; antibacterial; drug design.
 KW
 XX Haemophilus influenzae.
 OS
 XX
 PN WO200170955-A2.
 PD
 XX 27-SEP-2001.
 PF
 XX 21-MAR-2001; 2001WO-US09180.
 XX
 PR 21-MAR-2000; 2000US-191078P.
 PR 23-MAY-2000; 2000US-206848P.
 PR 26-MAY-2000; 2000US-207727P.
 PR 23-OCT-2000; 2000US-242578P.
 PR 27-NOV-2000; 2000US-253625P.
 PR 22-DEC-2000; 2000US-257931P.
 PR 16-FEB-2001; 2001US-269308P.
 XX
 PA (ELIT-) ELITRA PHARM INC.
 XX
 XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GT;
 PI Yamamoto RT, Xu HH;
 DR WPI; 2001-611495/70.
 DR N-FSDB; AAS53411.
 XX
 PS Example 3; Seq ID No 1145; 511pp; English.
 PT
 XX New polynucleotides for the identification and development of
 PT antibiotics, comprise sequences of antisense nucleic acids -
 XX
 XX The invention relates to antisense inhibitors of genes essential to
 CC prokaryotic cellular proliferation, their use in identifying the
 CC genes, their use in the discovery of novel antibiotics, the essential
 CC genes themselves and the encoded proteins. The prokaryotes used are
 CC *Baccherichia coli*, *Staphylococcus aureus*, *Salmonella typhi*, *Klebsiella*
 CC *pneumoniae*, *Pseudomonas aeruginosa* and *Enterococcus faecalis*. The
 CC invention is also useful for the identification of potential new targets
 CC for antibiotic development. The antisense nucleic acids can also be used
 CC to identify proteins used in proliferation, to express these proteins,
 CC and to obtain antibodies capable of binding to the expressed proteins.
 CC The proteins can be used to screen compounds in rational drug discovery
 CC programmes. The antisense nucleic acid sequence is also useful to screen

Db 850 GlySnaIaLeuAlaValThrPheAlaLeuAlaValIleIleIlePheLeuValLeuAla 869
 QY 1313 TTGAATGGCAATATGCTTTCAGTCCGTATGCGTTTAAATCATGACCTTTGGCTACC 1372
 Db 870 IlegInPheGluSerIleArgAspProMetValIleMetIleSerValProLeuAlaVal 889
 QY 1373 TGTGCAGCTTG-----TTTATGACACTTTCTTTGCAAGAAAATTCAAATA 1420
 Db 890 SerGlyAlaLeuValSerLeuAsnIleLeuSerPhePheSerIleAlaGlyThrThreu 909
 QY 1421 GATTGCAAGCCATTGGTGGCTTTATGACTGATGGGGTATTCTTAAACAATCTTGG 1480
 Db 910 AsnIleTySerGlnValGlyLeuIleThrLeuValGlyLeuIleThrLeuHisGlyLe 929
 QY 1481 ATCATTTTGTATCGTATTCGTGAAGATGCCAAGCAACTGTTTACCCCTATGCACTT 1540
 Db 930 LeuMetCysGluValAlaIleGluGlnLeuAsnHisGlyLysThrArgIleGluAla 949
 QY 1541 TTAGTTAATGATGCCCTTCAAAAGACGTTACGCCGACGATATGCAACAACATCAACT 1600
 Db 950 IleThrHis--AlaAlaIleValArgLeuArgProIleLeuMetThrThraIalaMet 968
 QY 1601 CTATCAGTTTGTATGCTTTCGTTTATGAGCGGCTCC-----TCTGCTTTTAT 1651
 Db 969 ValAlaGlyLeuIleProLeuLeuTyraIleThrGlyAlaValSerArgPheSer 988
 QY 1652 TTTCATTATTAATGACCAATAGGATCTTCTAGCACTTATGCTCTTATATTTGCA 1711
 Db 989 IleGlyIleValIleValAlaIleGlyLeuSerIleGlyThrIlePheThrLeuPheValLeu 1008
 QY 1712 CCA-----CCTGCTGTTGCTTATAGTCCGT 1738
 Db 1009 ProValAlaIleTySerTyraIleAlaThrGluHisLysProLeuProValPhe----- 1025
 QY 1739 AAGAAATCGCTCA 1753
 Db 1026 AspGluAsnLysThr 1030
 RESULT 41
 AAY19956
 ID AAY19956 standard; Protein; 1036 AA.
 AC AAY19956;
 XX
 DT 19-JUL-1999 (first entry)
 XX
 DE B. burgdorferi antigenic protein, f814.aa.
 XX
 KM Antigenic protein; vaccine; Lyme disease; infection; detection.
 XX
 OS Borrelia burgdorferi.
 XX
 PN MO859071-A1.
 PD 30-DEC-1998.
 XX
 PF 18-JUN-1998; 98WO-US12718.
 PR 03-SEP-1997; 97US-0057483.
 PR 20-JUN-1997; 97US-0050359.
 PR 22-JUL-1997; 97US-0053344.
 PR 22-JUL-1997; 97US-0053377.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 PA (MEDI-) MEDIMUNE INC.
 PT Choi GH, Erwin AL, Hanson MS, Lachigra R;
 WPI; 1999-189980/16.
 DR N-PSDB; AAX61653.
 XX
 PT New isolated Borrelia burgdorferi nucleic acids - used to develop
 products for the diagnosis, prevention and treatment of diseases

PT caused by Borrelia, particularly Lyme disease
 XX
 PS Claim 12; Page 135; 275pp; English.
 XX
 CC This sequence represents a Borrelia burgdorferi (Bb) protein of the
 CC invention, which is suitable for use in a vaccine. The Bb polypeptides
 CC can be used in vaccines for eliciting protective antibodies to members of
 CC the Borrelia genus, particularly for the use against Lyme disease in
 CC humans and animals. They can be used for preventing or attenuating an
 CC infection caused by a member of the Borrelia genus. The products can also
 CC be used for detection of members of the Borrelia genus.
 XX
 SQ Sequence 1036 AA.
 Alignment Scores:
 Pred. No.: 9.07e-06 Length: 1036
 Score: 136.50 Matches: 128
 Percent Similarity: 32.97% Conserved: 117
 Best Local Similarity: 17.23% Mismatches: 243
 Query Match: 4.12% Indels: 255
 DB: Gaps: 29
 US-09-868-987-1 (1-1864) x AAY19956 (1-1036)
 QY 173 GAATGAGCAAACTC-----GCCCTGATTTAAATCTGGAGGATGCTTTT 220
 Db 310 GlnIleGlnLysLeuLysLeuSerMetProLysAspMetLysLeu----- 324
 QY 221 GTTCCGAGGTTCTCACTGAGAGACAGCATCTCTGATCTTGGGAAAAACAATGTACA 280
 Db 325 -----GlnIleAlaSerasp-----SerThrAspHeIleLysAlaSerIleSer 339
 QY 281 CAAGCATTTATCTAGCATCTGCTTGGCAATCTTATGTTTGTGATGAGCGTATAT 340
 Db 340 ThrValAlaAsnSerAlaIleTyPheGlyAlaMetLeuAlaIlePhePhe 359
 QY 341 TATGAG---TTTGAGGCGTCATCGCTTGGGA---GCTGTTCTTGAAATCTTTCCTT 394
 Db 360 LeuArgSerPheArgAlaThrIleIleIleIleGlyIleSerIleProIleAlaIleValLeu 379
 QY 395 ATCTGGCAGCTCTACAGTATTTGGATGCCCACTGCTGTCAGGACTCGTGAGTT 454
 Db 380 ThrPheCysLeuMetTyrrPheValAsnIleSerLeuAsnIleMetSerLeuAlaGlyLeu 399
 QY 455 GTTCTGCTATGGGAGTGCCTGATGATCAAAAGTTCTTGATTTGCAAGATCCGAG 514
 Db 400 AlaLeuGlyIleGlyMetValAlaAspCysSerIleValAlaIleAspAsnIleTyrrLys 419
 QY 515 GAATTTTATTTGCTCAAGCTTMAAAANCTGTAAAGAAAGATATACCAAGGCTTTT 574
 Db 420 TyraArgLysGlyAlaLysLeuIleSerSerSerIleLeuGlyAlaGlnLysMet 439
 QY 575 GAGCCATTTTGTATCTTAATCTGACTGACCA-----GTATGGCTCAGACACTTTTTC 628
 Db 440 LeuProIleHisSerSerThrPheThrSerIleCysValPheGlyProPheLeuIlePhe 459
 QY 629 TTCCTAGATACAGGCGCT-----ATTAAGGTTTGCTTGACATTAATTTA 676
 Db 460 LysSerGlnLeuGlyAlaTyrrGlyAspPhePheLysAspPheThrPheThrIleValIle 479
 QY 677 GGAATTTTCTCTTCATGTTTAAGGCTCTTTCATGACTAAATTTTCTTCATGCTGTGG 736
 Db 480 SerLeuGlyValSerLeuLeuValAlaIlePheLeuValProValLeuSerHisTyrr 499
 QY 737 ATGAATAAGACCCAACTACATCAAGTTGATGATGATTAAGTTCTGAGG----- 787
 Db 500 ValGlyLeuTyrrThrSerPheGlnLysAsnIleLysAsnIleAlaGlyIleAsp 519
 QY 787 ----- 787
 Db 520 AlaPhePheAlaSerIleTyrrPheLeuGlnPheLeuTyrrIleAsnLeuLeuAsnIle 539
 QY 788 ---ATAAAGCATGATTTCTTGAGAGCATGCAAAAACATTTGGGCTGTTTCTGAAAGCTT 844

Db	894	SerPheLeuLysProPheIleIleIlePheThrIleProLeuThrAlaIleGlyValVal	913
Qy	1388	ATAGCACAATTTCTTTTGTG---AAGAAAATTCAAATAGATTTTGCAGCCATTGGTGCTTTA	1444
Db	914	LeuIleHisPheLeuAlaGlyGlyLysLeuSerIle---PheAlaAlaIleGlyMetLeu	932
Qy	1445	ATGACTGTATTGGGGTATTCAATTAACAATACTTTTGATCATCTTTTGATCGTATTTCGTGAA	1504
Db	933	MetLeuVal---GlyValValValAsnThrGlyIleValLeuValAsp---	947
Qy	1505	GATCGCCAGCGAACCTGTTTATCCCTCATGCATGTT-----	1540
Db	948	-----TyrThrGlyLeuLeuIleLysArgGlyPheGlyLeuArgGlu	961
Qy	1541	TTAGTTAATGATCCCTTCAAAAGACGTTTCAGCCACCGGTAATGACAAACAGCTACAACT	1600
Db	962	AlaIleIleGlySerCysArgSerArgLeuArgProIleLeuMetSerSerLeuThrSer	981
Qy	1601	CTATCAGTTTTGTTAAGCTTTTGTATTAGCGGCTCTCT------GTCTTTAAT	1651
Db	982	IleIleGlyLeuIleProMetAlaPheSerSerGlySerGlyAsnGluLeuLeuLysPro	1001
Qy	1652	TTTGCATTATTATGACCATAGGATTCTTCTAGGAACCTTTATCGTCTCTTTATATTGCA	1711
Db	1002	IleAlaPheThrPheIleGlyMetThrAlaSerThrPheLeuThrLeuPhePheIle	1021
Qy	1712	CCACCTCTG	1720
Db	1022	ProMetLeu	1024
RESULT	42		
ABP40160			
ID	ABP40160	standard; Protein; 1051 AA.	
XX	AC	ABP40160;	
XX	DT	24-JUL-2002 (first entry)	
XX	DE	Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:5005.	
XX	KW	Staphylococcus epidermidis; open reading frame; ORF; bacterial infection	
XX	KW	antibacterial; gene therapy.	
XX	OS	Staphylococcus epidermidis.	
XX	PN	US6380370-B1.	
XX	PD	30-APR-2002.	
XX	PF	13-AUG-1998; 98US-0134001.	
XX	PR	14-AUG-1997; 97US-055779P.	
XX	PR	08-NOV-1997; 97US-064964P.	
XX	PA	(GENO-) GENOME THERAPEUTICS CORP.	
XX	PI	Doucette-Stamm LA, Bush D;	
XX	DR	WPI; 2002-381255/41.	
XX	DR	N-PSDB; ABN92705.	
PT	PT	Novel isolated nucleic acid encoding a Staphylococcus epidermidis	
XX	PS	polypeptide, useful for diagnosing and treating bacterial infections -	
XX	PS	Disclosure; SEQ ID 5005; 267pp; English.	
CC	CC	ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading	
CC	CC	frame (ORF) nucleic acid sequences which encode the amino acid sequence	
CC	CC	given in ABP35124 to ABP37960. The S. epidermidis sequences have	
CC	CC	antibacterial activity and can be used in gene therapy. The sequences	
CC	CC	can also be used in the diagnosis and treatment of bacterial infection	
CC	CC	particularly S. epidermidis infections. The sequences can be used to	
CC	CC	screen for compounds able to interfere with the S. epidermidis life	

CC cycle or inhibit S. epidermidis infection.
 CC N.B. The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from the
 CC USPTO web site.

XX Sequence 1051 AA;

Alignment Scores:

Pred. No.:	9.14e-36	Length:	1051
Score:	136.50	Matches:	147
Percent Similarity:	34.12%	Conservative:	127
Best Local Similarity:	18.31%	Mismatches:	266
Query Match:	4.12%	Indels:	263
DB:	23	Gaps:	32

US-09-868-987-1 (1-1864) x ABP40160 (1-1051)

```

QY 56 TCTGCAACGGTGGATGGCGTATGAGTTCAGTTCATATGTCAGACGCCCT 115
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 273 SerThrAsnSerSerThrHisMetAlaGln-----LysGlyGlnMetProser----- 288

QY 116 ATTTAAACGTCCCATGAAATATCCAGTGTCTCAGGAAATTTACCCACCGTAA 175
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 289 -----ValProLeuLysAspLeuAlaAspIleSerValGlyAspGluArgThrSer 305

QY 176 GTGAGCAA----- 184
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 306 IleSerLysThrAsnGlyLysAspAlaValAsnLeuGlnIleMetLysSerGlnAspAla 325

QY 185 -----CTCGCGCTCAGATTTCGAGGAGTGTCTTGTGTCGCGAGGT 232
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 326 AsnThrValGlnValAlaAspGluValGlnLysLysValAspGluPheValArgAsnGlu 345

QY 233 CTCAGTAGAGACGATCTTCTT---GATCTTGAGAA----- 268
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 346 SerGlyMetLysSerIleLysThrMetAspThrAlaLysProIleGluAspSerLeuTyr 365

QY 269 -----AAACAATGATCAACAAGCATTTCTCAGCATCTGTGGCTTGCAATGCTT 319
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 366 ThrMetValGluLysAlaAlaLeuGlyThrIle-----ValAlaIleIle 380

QY 320 ATTGTTTGTGATGAGCGTATTTATAGATTGGAGCGTCACTGCGGAGCTGTTCTT 379
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 381 ValIleLeuLeuPheLeuAlaAsnIleArgThrThrAlaIleSerIleValSerIlePro 400

QY 380 CTGATATCTTTGCTTATCTGGCAGCTCTACAGATTGTCAGCCACTGACCTTGCA 439
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 401 MetSerIleLeuIleAlaLeuIleAlaLeuLysLeuSerAsnValSerLeuAsnIleLeu 420

QY 440 GGACTCGCTGGAGATTGCTCTGATGGGATGGCCGATGATGCAATGTTCTGTATTC 499
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 421 ThrLeuGlyAlaLeuThrValAlaIleGlyArgValIleAspAspSerIleValValVal 440

QY 500 GAAAGATTC-----CGAGAGAAATTTTATGTCGCAAGCTTAA 541
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 441 GluAsnIlePheArgLysLeuSerAspProAsnGluLysLysGlyGluAsnIleLeu 460

QY 542 AAATCTGTAGAAAAAGATATACCAAGCTTTTGAGAGCACTTTTGATTTACTGACT 601
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 461 IleSerIleAlaThrArg-----GluValPheLysProIleMetSerSerThrLeuVal 477

QY 602 ACAGTATTTGGCTCAGACCTTTCTTCTCTAGATACAGGCGCTATT----- 649
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 478 ThrIleValValPheLeuProLeuValPheVal---SerGlySerValGlyGluMetPhe 496

QY 650 AAAGGCTTGTGTCATGATTTAGGATTTTCTTCAAGTTTAAAGCTCTTTTC 709
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 497 ArgProPheAlaLeuAlaIleThrPheSerLeuLeuAlaSerLeuValSerIleThr 516

QY 710 ATG-----ACTAAATTTTCTTCATGCTGTGATGATGATTAAGACCAACAT 754
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 517 LeuValProSerLeuGlyAlaThrPhePheLysAsnGlyValLysAsnArgGluGlnLys 536
  
```

```

QY 755 ACACAGTGCATATATG----- 772
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 537 GluGlyLeuGlyThrValGlyAlaGlyAlaTyrArgSerAlaLeuAsnTrpSerLeuAsnHis 556

QY 772 ----- 772

Db 557 LysTrpIleValLeuIleValSerIlePheIleLeuValGlySerValIleGlyAla 576

QY 773 ----- 796
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 577 ArgAsnLeuGlyThrSerTyrIleSerThrGlyAspAsnLysPheLeuAlaLeuThrTyr 596

QY 797 ----- 799
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 597 ThrProLysProGlyLysThrGlnLysSerValThrGlnHisAlaGluLysValGlnAsn 616

QY 800 TTCTTGAGAGATGACAAAACCTTTGGCGTGT----- 832
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 617 TyrLeuAspLysLysAspLysValGluThrValGlnTyrSerIleGlyGlyProThrPro 636

QY 833 -----CTCGAAGTGTCTTTTCTTTAGCTTGGCTGCT----- 865
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 637 GlnAspProThrGlySerThrAsnSerMetAlaIleMetIleLysTyrGlnSerAspThr 656

QY 865 ----- 865

Db 657 ProAsnPheAspGluGluProAspLysValLeuLysHisIleGluThrPheLysGlnPro 676

QY 866 -----CTCGGCTTTGGAGCTGCAATTCGTTTGGGATGATGAT 904
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 677 GlyGlyTyrLysAsnGlnAspLeuGlyThrGlyAlaGlyAsnAsnSerValGluValThr 696

QY 905 TTTAAAGCA-----GGTATGCTTATCTTAACTTTCATCAAAAGCATGGCATC 932
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 697 ValLysGlyProAsnThrSerAlaMetLysAspThrValAsnArgValGluLysMetMet 716

QY 953 AGCGATGTT----- 961
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 717 ThrAspIleLysGlyIleThrAsnValLysSerAspLeuSerGlnThrTyrAspGlnTyr 736

QY 962 -----GCTCAATGCGT 973
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 737 GluIleLysValAspGlnAsnLysAlaAlaAspAsnGlyIleSerAlaIleGlnLeuAla 756

QY 974 GGCAAAGTGTGCATTAACATACAGAGAGCTGCTTCT----- 1012
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 757 MetAsnLeuAsnGlnLysLeuProGluLysThrIleSerThrValAsnGluLysGlyLys 776

QY 1013 TCTAGAGACTTCGCTATT-----CAACAATTTGATCTTCAGAAAAGATCAAA 1060
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 777 SerIleAspValLysValLysGlnAsnLysGlnThrAspTrpSerSerGlnLysIleLys 796

QY 1061 ATCTAATTTAGTAAAGCTTTAAGCTTATCTAAGCAGATACGA-----GCCCTGTC 1114
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 797 AsnIleLysLeuAsnLysProThrGlyGlyThrIleLysLeuSerGluIleAlaSerLeu 816

QY 1115 -----CTAAATTTAAGATATGAGCTGCGGCTTATGTTGGGATGTTT 1156
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 817 LysLysSerTyrThrProSerLysLeuThrGlnGluAspGlyAspTrpAlaThrThrVal 836

QY 1157 GTC---AGAAACAGGCGTATGTTCTCTACGGAAGCTTAAAGCAAAATTTTGG 1213
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 837 ThrGlyLysValAlaThrAspLysAspValGlyGlyLysSerGln-----GlnValMet 853

QY 1214 TCAAAGGTAAAGCAAACTATGCAAAATTCGTTATTCAGGACATCGG----- 1267
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 854 AlaLysValLysAsp---LeuGluLysProSerHisIleLysLysAsnValGlyGlyAla 872

QY 1268 -----CTTTAGAGCTTTGGCA 1285
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 873 ThrAspAspIleAspLysAlaIleSerGlnLeuAlaMetIleAlaIleIle 892

QY 1286 ATCATCTTGCTATGATGAGTTGGCGCTTGAATGCAATATGCT---TTCACTGCGCGTA 1342
  
```


Db 893 ileValThrLeuileLeuValileThrPheArgGlyLeuAlaProPheThrIleLeu 912
Qy 1343 TCGCTTTAAATTCATGACCTTTGGCTACCTGTCAGCTGTTGTTATAGCACATTCTTT 1402
Db 913 PheSerLeuProPheThrValileGlyValValLeuAlaLeuileIleThrGly----- 930
Qy 1403 TTGAAGAAAATTCAAATAGATTGCAAGCCATTGCTGTTTAAATGACTGTATTGGGGTAT 1462
Db 931 ---GluThrIleSerValPro---SerLeuileGlyMetLeuMet---LeuileGlyIle 947
Qy 1463 TCATTAAACAATACCTTTCATCATTTTTCATCGTATTTCGTAAGATCGCAAGCAACCTG 1522
Db 948 ValValThrAsnAlaileValleuileaspargvalileAsnAsnGluGlnGlyMet 967
Qy 1523 TTTACCCCTATGTCATGTTTATGTTAATGATGCCCTTCAAAAGACGTTTCAGCGCACGGTA 1582
Db 968 -----PrometLysGlnAlaLeuileGluAlaGlyThrArgIleArgProIleLeu 985
Qy 1583 ATGACACAGCTACACTCTATCAGTTTGTAAATGCTTTTGTATAGCGGCTCTCT 1642
Db 986 MetThrAlaileAlaThrIleGlyAlaLeuValProLeuLeuPheGlyHisAspSerSer 1005
Qy 1643 GTCITT-----AATTGTCATTATTATGACCATAGGATTCCTTAGGACTTTA 1693
Db 1006 IleLeuileSerLysGlyMetAlaAlaThrValileGlyLeuileSerSerThrLeu 1025
Qy 1694 TCGTCTCTTTTATGACCACTCTG-----TTGTTGTTTATGCTCGCTAAAGAAAT 1747
Db 1026 LeuThrLeuPheValProValileThrGluIleLeuPheThrLeuLysAsnLysIle 1045
Qy 1748 CGCTCAAAA 1756
Db 1046 ThrLysLys 1048
RESULT 43
AAG90442
ID AAG90442 standard; Protein; 730 AA.
XX
AC AAG90442;
XX
XX 26-SEP-2001 (first entry)
XX
XX C glutamicum protein fragment SEQ ID NO: 4196.
DE
XX Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
XX organic acid synthesis.
XX
XX Corynebacterium glutamicum.
XX
XX EP1108790-A2.
XX
XX 20-JUN-2001.
XX
XX 18-DEC-2000; 2000EP-0127688.
XX
XX 16-DEC-1999; 99JP-0377484.
XX 07-APR-2000; 2000JP-0159162.
XX 03-AUG-2000; 2000JP-0280988.
XX
XX (KYOW) KYOWA HAKKO KOGYO KK.
XX
XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
XX Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX
XX WPI; 2001-376931/40.
XX N-PSDB; AAH65661.
XX
XX Novel polynucleotides derived from Coryneform bacteria, for identifying
XX mutation point of a gene, measuring expression of a gene, analysing
XX expression profile or pattern of a gene and identifying homologous gene

PS Claim 17; SEQ ID NO: 4196; 246pp + Sequence Listing; English.
XX
CC The present invention provides a number of nucleotide and protein
CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These
CC are useful for identifying the mutation point of a gene derived from a
CC mutant of coryneform bacterium, measuring expression amount and
CC analysing the expression profile or expression pattern of a gene derived
CC from Coryneform bacterium, and identifying a homologue of a gene derived
CC from coryneform bacterium. Coryneform bacteria are useful for producing
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC particularly L-lysine. The present sequence is a protein described
CC in the exemplification of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC European Patent Office.
XX
SQ Sequence 730 AA;
Alignment Scores:
Pred. No.: 3.6e-05 Length: 730
Score: 130.50 Matches: 128
Percent Similarity: 33.55% Conservative: 78
Best Local Similarity: 20.85% Mismatches: 270
Query Match: 3.94% Indels: 138
DB: 22 Gaps: 23
US-09-868-987-1 (1-1864) x AAG90442 (1-730)
Qy 125 GTCCATTGAAAAATCATGCCAGTGTCTCAGGGAATTTACCCACCGTGAAGTCAGCAA 184
Db 132 ValProLeuAsnProSerAlaGluLeuThr-----GluSerValGluLys 146
Qy 185 CTCGCCTCAGATTTAAAAATCTGGAGCGATGCTTTTGTTCCTCCGAG---GTTCTCAGTGAA 241
Db 147 LeuSerGluThrLeuThrGlnThrProAspThrValSerThrValThrGlyPro 166
Qy 242 GAGACGATCTCTCTGATCTTTGGGAAAAACAATGT---ACACAAGGCAATTTCTCAGCA 298
Db 167 AlaGlyPheThrAlaAspLeuSerAlaAlaPheAlaGlyIleAspGlyLeuLeuAla 186
Qy 299 TGCTGTGGCTGGCAATGCTTATTGTTTGTATGAGCGTATATTATAGATTT----- 349
Db 187 ValAlaLeuAlaAlaValLeuValileValileValThrValThrArgSerPheIleLeuPro 206
Qy 350 GCAGCGTCATCGCTTCGGGA-----GCTGTTCTTGAATCTTTGCTTATCTGGCA 403
Db 207 IleAlaValLeuAlaThrSerLeuPheAlaLeuThrValAlaLeuLeuValThrTrp 226
Qy 404 GCTCTACAGTATTGGATGCGCCACTCACCTTGTGAGACTCGCTGGGATT-----GTT 457
Db 227 LeuAlaLysTrp---AspIleLeuLeuLeuSerGlyGlnThrGlnGlyIleLeuPheIle 245
Qy 458 CTTGCTATGGGATGGCGGTAGATGCAATGTTCTTGTATTTCGAAGATCCGAGAGAA 517
Db 246 LeuValileGlyAlaAlaThrAspThrSerLeuLeuThrValAlaAlaArgPheArgGlu 265
Qy 518 TTTTATTGTCTCAAGCTTTAAAAAATCTGTAGAAAAAGGATATACCAGGCTTTTGA 577
Db 266 LeuArgValGlnGlnAspLysGlyIleAlaThrGlyLysAlaIleArgAlaSerValGlu 285
Qy 578 GCCATTTTGTGATCTTAACACTACAGTATTGGCTCAGCACTTTCTTTCTCTAGAT 637
Db 286 ProIleLeuAlaSerGlySerThrValileAlaGlyLeuLeuCysLeuLeuPheSerAsp 305
Qy 638 -----ACAGGCGCTATTAAAGGTTTGTGACATTGATTAGGA 679
Db 306 LeuLysSerAsnSerThrLeuGlyProValAlaSer-----ValGly 319
Qy 680 ATTTCTCTTCAATGTTTACGGCTCTT---TTCATGACTAAATTTTCTTCATG----- 730
Db 320 IleIlePheAlaMetLeuSerAlaLeuThrLeuLeuProAlaLeuLeuPheValPheGly 339
Qy 730 ----- 730

Db 340 ArgValAlaPheTrpProlyValArgProlyGluProGluValAlaArgAlaIysAsn 359
 QY 731 -----CTGTGATGAATAGACCCACATACACAGTTGATGATG 772
 Db 360 AspIleProAlaSerGlyIleTrpSerLysValAlaAspLeuValGluGlnHisProArg 379
 QY 773 AATAAGTTCGTGGGATAAAGCATGATTTCTGAGAGAGGCAAA----- 817
 Db 380 AlaIleTrpValSerThrLeuIleValLeuLeuGlyAlaAlaPheValProThrLeu 399
 QY 818 AAACCTTTGGGCTGTTGAGAGTCTTTCTTTAGTTCGCTCGGGTTGA 877
 Db 400 LysAlaAspGlyValSerGlnSerAspLeuValLeuGlySerSerGluAlaArgAspGly 419
 QY 878 GCCTGGAATTCGTTTGGAGATGATTTAAAGAGAGG-----TATGCC 922
 Db 420 -----GlnGlnAlaLeuGlyGluHisPheProGlyGlySerGlySerProAlaTyrIle 437
 QY 923 TTAACTTTAATCCAAAAGACATGCATGCATGTTGCTCAAAATGCGTGCAGAAAGTT 982
 Db 438 IleValAspGluThrGlnAlaAlaGlnAlaAlaAspValVal----- 451
 QY 983 GTGCATAACTACAGAGAGCGCTTTCTTCTAGAGCTTCGATTCGATCAACATTTGA 1042
 Db 452 -----LeuAsnAsnAspAsnPheGluThrValThrValThr 463
 QY 1043 TCTTCAGAAAAGATCAAAATCTATTATTAGATTAAGCTTTAAGCTTACTAACAGATA 1102
 Db 464 SerAlaAspSer-ProSerGlySerAlaProIleThrAlaAspGlyIleVal----- 480
 QY 1103 CGAGCTCTCTCTCTAAATTAACATCATCATGCTGGCTTATTGGGATTTGTTCAGA 1162
 Db 481 -----ProLeuGly-----SerGlyThrAlaProGlyProValValVa 493
 QY 1163 AACAGCGCTGATTTCTCTACGAAACTCTA-----AACGAAAGCA 1204
 Db 493 IGluGlyGlnValLeuLeuGlnAlaThrLeuValGluAlaProAspSerGluGluAlaGly 513
 QY 1205 AAATTTGGTCAAGATGACGAGCAAACTATCGAAGAAATGCTTATACGCGCAGCATC 1264
 Db 513 nLys-AlaIleArgSerIleArgGlnThrPheAlaAspGluAsnIleSer-AlaValVal 532
 QY 1265 GGGCTTTTAAAGCTTTGGCAATCATCTGCTATGTAGTTCGCTTGAATGCGCA 1324
 Db 533 GlyGlyValThrAlaThrSerVal----- 540
 QY 1325 TATGCTTCAAGTCCGATGCGCTTAAATTCATGACCT-----TTGGCT 1369
 Db 541 -----AspThrAsnAspAlaSerIleHisAspArgAsnLeuIleIleProIleVal 557
 QY 1370 ACCTGTCAGCTCTGTTTACGACATTTCTTTGAAGAAATTCAAATGATTTGCA 1429
 Db 558 LeuLeuValIleLeuValIleLeuMetLeuLeuArgSerIleValAlaProLeuLeu 577
 QY 1430 GCCATTGGTCTTAAATGATGTTATGGGATTCATTAAACAATCTTATCATTTT 1489
 Db 578 LeuValValThrThrValValSerPheAlaThrAlaLeuGlyValAlaAlaLeuLeuPhe 597
 QY 1490 GATGTAATTCCTGAAGATCGCCAGCGCAACGTTTACCCTATGCACTTTAGTTAGTT 1546
 Db 598 AsnHisValPheSerPheProGlyAlaAspProAlaValProLeuTyrGlyPheValPhe 617
 QY 1547 -----AATGATGCCCTTCAAAAAGAGTTCAGCGCAGCGTA 1582
 Db 618 LeuValAlaLeuGlyIleAspTyrAsnIlePheLeuValThrArgIleArgGluGluThr 637
 QY 1583 ATGACAAACAGCTACACTATACAGTTTGTAAATGCTTTTGTATAGCGGCTCTCT 1642
 Db 638 LysThrHisGlyThrArgLeuGlyIleLeuArgGlyLeuThrValThrGlyGly----- 655
 QY 1643 GTCTTAATTTTGCATTATATAGCAACCTAGGAAATCTTCA-----GGAACCTTATTCGCT 1699

Db 656 -----ValIleThrSerAlaGlyValValLeuAlaAlaThrPheAlaAla 670
 QY 1700 CTTTATATTGCACACCTCTCTGTTGTTTATGCTC 1735
 Db 671 LeuTyrValIleProIleLeuPheLeuAlaGlnIle 682
 RESULT 44
 AAB78979
 ID AAB78979 standard; Protein; 730 AA.
 AC AAB78979;
 XX
 XX
 DT 30-APR-2001 (first entry)
 XX
 DE C. glutamicum SRT protein sequence SEQ ID NO:218.
 XX
 XX Corynebacterium glutamicum; stress; resistance; tolerance; SRT;
 KW fine chemical production; organic acid; proteino-genic amino acid;
 KW nonproteino-genic amino acid; purine base; pyrimidine base; nucleoside;
 KW nucleotide; lipid; saturated fatty acid; unsaturated fatty acid; diol;
 KW carboxylate; aromatic compound; vitamin; cofactor; polyketide; enzyme;
 KW evolutionary study; environmental hazard; fermentation.
 XX
 OS Corynebacterium glutamicum.
 XX
 PN MO200100804-A2.
 XX
 PD 04-JAN-2001.
 XX
 PF 23-JUN-2000; 2000MO-IB00922.
 XX
 PR 25-JUN-1999; 99US-0141031.
 PR 01-JUL-1999; 99DE-1030429.
 PR 01-JUL-1999; 99US-0142692.
 PR 08-JUL-1999; 99DE-1031413.
 PR 08-JUL-1999; 99DE-1031457.
 PR 08-JUL-1999; 99DE-1031541.
 PR 09-JUL-1999; 99DE-1032209.
 PR 09-JUL-1999; 99DE-1032230.
 PR 14-JUL-1999; 99DE-1032914.
 PR 27-AUG-1999; 99DE-1040764.
 PR 27-AUG-1999; 99US-0151214.
 PR 31-AUG-1999; 99DE-1041382.
 XX
 PA (BAD) BASE AG.
 XX
 PI Pompeius M, Kroeger B, Schroeder H, Zelder O, Haberhauer G, Lee H,
 PI Kim H;
 XX
 DR WPI; 2001-061972/07.
 XX
 XX N-PSDB; AAF71092.
 XX
 PT New isolated Corynebacterium glutamicum nucleic acid encoding a stress,
 PT tolerance or resistance protein, for production or modulation of
 PT production of fine chemicals, such as, e.g. amino acids, lipids,
 PT carbohydrates, or enzymes -
 XX
 PS Claim 20; Page 397-399; 526pp; English.
 XX
 XX AAF7984 to AAF7113 encode the Corynebacterium glutamicum stress,
 CC resistance, and tolerance (SRT) proteins given in AAB78971 to AAB79020.
 CC The C. glutamicum SRT genes (I) can be used in vectors (II) for
 CC expression in host cells and production of fine chemicals, such as, an
 CC organic acid, a proteino-genic or nonproteino-genic amino acid (preferred),
 CC a purine or pyrimidine base, a nucleoside, a nucleotide, a lipid, a
 CC saturated or unsaturated fatty acid, a diol, a carboxylate, an aromatic
 CC compound, a vitamin, a cofactor, a polyketide, or an enzyme. The fine
 CC chemical production can be modulated. The presence of (I) or the SRT
 CC proteins (III) encoded by them are used for diagnosing the presence
 CC or activity of Corynebacterium diptheriae. (I), (II), (III) and host
 CC cells containing them can be used to map the genomes of organisms related
 CC to C. glutamicum, to identify and localise C. glutamicum sequences of
 CC interest, in evolutionary studies, in determination of SRT protein

CC regions required for function, in modulating the SPT protein activity,
 CC and in modulating the activity of an SRP pathway. (ii) are used to permit
 CC C. glutamicum to survive in an environment that is normally
 CC environmentally or chemically hazardous to it. (i) and protein molecules
 CC encoded by it increase the survival of C. glutamicum to chemical and
 CC environmental hazards and provide a means for continued growth and
 CC multiplication in large scale fermentative growth conditions. By
 CC increasing the growth rate or maintaining a normal growth rate in poor or
 CC toxic conditions, the yield, production and/or efficiency or production
 CC of fine chemicals from a culture may be increased.

XX Sequence 730 AA;

Alignment Scores:

Pred. No.:	3.6e-05	Length:	730
Score:	130.50	Matches:	128
Percent Similarity:	33.55%	Conservative:	78
Best Local Similarity:	20.85%	Mismatches:	270
Query Match:	3.94%	Indels:	138
DB:	22	Gaps:	23

US-09-868-987-1 (1-1864) x AAB78979 (1-730)

Qy	125	GTCCATTGAAAAATCATGCCAGTCTCTCAGGAAATTTACCCCGTGAAGTGAGCAA	184
Db	132	ValProLeuAsnProSerAlaGluLeuThr-----GluSerValGluLys	146
Qy	185	CTCGCCTCAGATTTAAATCTCGAGCGATGCTTTTTCCTCCGAG---GTTCTCAGTGAA	241
Db	147	LeuSerGluThrLeuThrGlnGlnThrProAspTyrValSerThrTyrValThrGlyPro	166
Qy	242	GAGACGATCTCTTCGTATCTGGGAAAAACAATGT---ACACAAGGCATTATCTCAGCA	298
Db	167	AlaGlyPheThrAlaAspLeuSerAlaAlaPheAlaGlyIleAspGlyLeuLeuAla	186
Qy	299	TGCTGCGCTTGGCAATCTTATCTTTGATGACGCTATATATAGATT	349
Db	187	ValAlaLeuAlaValLeuValIleLeuValIleValTyrArgSerPheIleLeuPro	206
Qy	350	GGAGCGCTATCGCTTCGGGA-----GCTGTTCTTCTGAATCTTTTCTTATCTGGCA	403
Db	207	IleAlaValLeuAlaThrSerLeuPheAlaLeuThrValAlaLeuLeuValThrTrp	226
Qy	404	GCTTACAGTATTTGGATCGCCACTACCTGTGTCAGGACTCGCTGGGATT-----GTT	457
Db	227	LeuAlaLysTrp---AspIleLeuLeuSerGlyGlnThrGlnGlyIleLeuPheIle	245
Qy	458	CTTGCTATGGGATGCGGTAGATCAATGTTCTTGTATTGGAAGAAATCCGAGAGAA	517
Db	246	LeuValIleGlyAlaAlaThrAspTyrSerLeuLeuTyrValAlaArgPheArgGluGlu	265
Qy	518	TTTTTATTGCTCTCAAGTCTTAAAAAATCTGTAGAAAAAGGATATACCAAGCTTTTGA	577
Db	266	LeuArgValGlnAsnAspGlyIleAlaThrGlyLysAlaIleArgAlaSerValGlu	285
Qy	578	GCCATTTTGTATCTTAATGACTACAGTATGGCCCTCAGCACTTCTTTCTCTAGAT	637
Db	286	ProIleLeuAlaSerGlySerThrValIleAlaGlyLeuLeuCysLeuLeuPheSerAsp	305
Qy	638	-----ACAGGCTTATTAAAGGTTTGTTCACATTGATTTTAGA	679
Db	306	LeuLysSerAsnSerThrLeuGlyProValAlaSer-----ValGly	319
Qy	680	ATTTTCTCTCAATTTTACGCTCTT---TTCATGACTAAATTTTCTTCATG-----	730
Db	320	IleIlePheAlaMetLeuSerAlaLeuThrLeuLeuProAlaLeuLeuPheValPheGly	339
Qy	730	-----	730
Db	340	ArgValAlaPheTrpProLysArgProLysTyrGluProGluLysAlaArgAlaLysAsn	359
Qy	731	-----CTGGATGATTAAGACCCCAACATACACATTCGATATGATG	772

Db	360	AspIleProAlaSerGlyIleTrpSerLysValAlaAlaAspLeuValGluGlnHisProArg	379
Qy	773	ANTAAGTTCTGGGATAAAGCATGATTCTTGTAGAGGATGCAAA-----	817
Db	380	AlaIleTrpValSerThrLeuIleValLeuLeuGlyAlaAlaPheValProThrLeu	399
Qy	818	AACTTTTGGGCTGTTCTGGAAGTGTCTTTCTTTTGTGCTGCTCTCGGTTTGGGA	877
Db	400	LysAlaAspGlyValSerGlnSerAspLeuValLeuGlySerSerGluAlaArgAspGly	419
Qy	878	GCCTGGAATTCGTTTGGGAATGGATTTTAAAGGAGGG-----TATGCC	922
Db	420	-----GlnGlnAlaLeuGlyGluHisPheProGlyGlySerGlySerProAlaTyrIle	437
Qy	923	TTTACCTTTAATCAAAAGAGCATGGCATCAGCGATCTTCTCAAAATCGGTGGCAAGTT	982
Db	438	IleValAspGluThrGlnAlaAlaGlnAlaAlaAspValVal-----	451
Qy	983	GTGCATAACTACAGGAAGCTGCTTTCTTCTAGACACTTCGCTATTCAACATTTGGA	1042
Db	452	-----LeuAsnAsnAspAsnPheGluThrValThrValThr	463
Qy	1043	TCTTCAGAAAAGATCAAAATCTATTTTAGTGATAAAGCTTTAAGCTATATCAAGCAGATA	1102
Db	464	SerAlaAspSer-ProSerGlySerAlaProIleThrAlaAspGlyIleVal-----	480
Qy	1103	CGAGCCTCTCTCTCAAAATTAACGATCATGAGCTGGCGTTTATGTGGGATTTGTTCAGA	1162
Db	481	-----ProLeuGly-----SerGlyThrAlaProGlyProValVal	493
Qy	1163	AACAGGCTAGATTCTCTACGGAAACTCTA-----AACGAAACGCA	1204
Db	493	IGluGlyGlnValLeuLeuGlnAlaThrLeuValIGluAlaProAspSerGluGluAlaG	513
Qy	1205	AAATTTTGGTCAAGGTAAAGCAGCAAACTATCGAAGAAATTCGCTTATCAGGCCACCATC	1264
Db	513	nLys-AlaIleArgSerIleArgGlnThrPheAlaAspGluAsnIleSer-AlaVal	532
Qy	1265	GGGTTTTAGAGCTTTGGCAATCATCTGCTCTATGTGAGTTTTCGCTTTGAATGGCAA	1324
Db	533	GlyGlyValThrAlaThrSerVal-----	540
Qy	1325	TATGCTTTTCAAGTCCGATGCGCTTTAATTCATGACCTT-----TTGGCT	1369
Db	541	-----AspThrAsnAspAlaSerIleHisAspArgAsnLeuIleIleProIleVal	557
Qy	1370	ACCTGTGCGTCTGTTTATAGACACATTTCTTTTGAAGAAATTCAAATAGATTTGCAA	1429
Db	558	LeuLeuValIleLeuValIleLeuMetLeuLeuLeuArgSerIleValAlaProLeuLeu	577
Qy	1430	GCCATTGCTGCTTTAATGACTGTATTGGGGTATTTCATTAACAATACTTTTGATCATTTT	1489
Db	578	LeuValValThrThrValValSerPheAlaThrAlaLeuGlyValAlaAlaLeuLeuPhe	597
Qy	1490	GATCGTATTCGTGAAGATCGCCAAAGCAACCTGTTTACCCTATGCATGTTTATGTT---	1546
Db	598	AsnHisValPheSerPheProGlyAlaAspProAlaValProLeuTyrGlyPheValPhe	617
Qy	1547	-----AATGATGCCCTTCAAAAGACGTTTCAGCCGCGCAGCGTA	1582
Db	618	LeuValAlaLeuGlyIleAspTyrAsnIlePheLeuValThrArgIleArgGluGluThr	637
Qy	1583	ATGACACAGCTACAACTCTATCAGTTTGTAAATGCTTTTGTATAGGGCGGCTCTCT	1642
Db	638	LysThrHisGlyThrArgLeuGlyIleLeuArgGlyLeuThrValThrGlyGly-----	655
Qy	1643	GTCTTTAATTTTGCATTTATATGACCATAGGATCTTCTTA---GGAACCTTTATCGCT	1699
Db	656	-----ValIleThrSerAlaGlyValValLeuAlaAlaThrPheAlaAla	670
Qy	1700	CTTTATATTGACCACTCTGTTGTTGTTTATGTC	1735
Db	671	LeuTyrValIlePheLeuPheLeuAlaGlnIle	682

Search completed: December 17, 2002, 12:35:33
Job time : 103.5 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model
Run on: December 17, 2002, 12:30:02 ; Search time 57.5 Seconds
(without alignments)
13359.020 Million cell updates/sec

Title: US-09-868-987-1
Perfect score: 3312
Sequence: 1 agggacttcgcatattgtc.....atgacaaattcgataatgc 1864

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 671590 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 1343160

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:
-MODEL=frame+ n2p.model -DEV=xlh
-DB=SPTRMBL 21 -QFMT=fastan -SUFFIX=rspt -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=45 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEA=SIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USRR=US09868987 @CGN 1 1 70 @runat 16122002 083006 23540 -NCPU=6 -ICPU=3
-NO XLPXY -NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -LONGLOG -DEV TIMEOUT=120
-WEN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_rvirus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2874	86.8	1402	16 Q927Y9	Q927Y9 chlamydia p

2	2083	62.9	1400	16	084454
3	2088	62.4	1400	16	084454
4	840	25.4	876	16	084454
5	790.5	23.9	849	16	084454
6	715.5	21.6	845	16	084454
7	685	20.7	759	16	084454
8	682	20.6	858	16	084454
9	682.5	20.6	844	16	084454
10	678.5	20.5	737	16	084454
11	668.5	20.2	768	16	084454
12	653.5	19.7	775	16	084454
13	630	19.0	795	16	084454
14	620.5	18.7	754	16	084454
15	618.5	18.7	754	16	084454
16	522.5	15.8	873	16	084454
17	460.5	13.9	411	16	084454
18	436	13.2	474	16	084454
19	424	12.8	620	16	084454
20	423.5	12.8	419	16	084454
21	403.5	12.2	526	16	084454
22	400.5	12.1	518	2	084454
23	399.5	12.1	616	16	084454
24	395.5	11.9	518	16	084454
25	391	11.8	417	16	084454
26	381	11.8	532	16	084454
27	390	11.8	323	16	084454
28	384.5	11.6	322	16	084454
29	383.5	11.6	615	16	084454
30	381.5	11.5	410	16	084454
31	376	11.4	617	16	084454
32	373	11.3	626	16	084454
33	369.5	11.2	306	16	084454
34	368	11.1	612	16	084454
35	365.5	11.0	437	16	084454
36	365	11.0	323	16	084454
37	364	11.0	618	16	084454
38	364	11.0	618	16	084454
39	363.5	11.0	315	16	084454
40	357.5	10.8	614	16	084454
41	357.5	10.8	615	2	084454
42	356.5	10.8	615	16	084454
43	356.5	10.8	615	16	084454
44	356	10.7	615	2	084454
45	355.5	10.7	465	16	084454

ALIGNMENTS

RESULT 1

ID	Q927Y9	PRELIMINARY;	PRT;	1402	AA.
AC	Q927Y9;				
DT	01-MAY-1999	(TREMBLrel. 10, Created)			
DT	01-MAY-1999	(TREMBLrel. 10, Last sequence update)			
DT	01-JUN-2002	(TREMBLrel. 21, Last annotation update)			
DE	Protein export proteins SECDF/SECF (Fusion) (SECDF protein, putative).				
DE	SECDF/SECF OR SECDF OR SECDF OR CPN0564 OR CP0186.				
GN	Chlamydia pneumoniae (Chlamydia pneumoniae).				
OS	Chlamydia pneumoniae (Chlamydia pneumoniae).				
OC	Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.				
OX	NCBI_TaxID=83558;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	MEDLINE=99206606; PubMed=10192388;				
RX	Kaiman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W., Olinger L., Grimwood J., Davis R.W., Stephens R.S.; "Comparative genomes of Chlamydia pneumoniae and C. trachomatis."; Nat. Genet. 21:385-389(1999).				
RL	[2]				
RN	SEQUENCE FROM N.A.				
RP	MEDLINE=20150255; PubMed=10684935;				
RX	MEDLINE=20150255; PubMed=10684935;				

QY 1559 CAAAGACGTTACGCCGACCGTAATGACACAGCCTACAACTCTATCAGTTTCTTAATG 1618
 Db 1337 GlnYstnRPhSeSerArgThrValMetThrThrAlaThrThrLeuSerValLeuMet 1356
 QY 1619 CTTTGTGTTATAGCGGCTCTCTGTCTTTAAATTTTGCATTTATATGACCATAGGAGTT 1678
 Db 1357 LeuLeuPheIleGlySerSerValPheAsnPheAlaPheIleMetThrIleGlyIle 1376
 QY 1679 CTTCTAGAACTTATCTCTCTTTATATGACACCTCTGTGTTGTTTATGTCGGT 1738
 Db 1377 LeuLeuGlyThrLeuSerSerLeuTyrIleAlaProProLeuLeuLeuPheMetValArg 1396
 QY 1739 AAGAAATCCCTCAAAA 1756
 Db 1397 LysGluAsnArgSerLys 1402
 RESULT 2
 ID 084454 PRELIMINARY; PRT; 1400 AA.
 AC 084454, 1998 (TrEMBLrel. 08, Created)
 DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Protein export.
 DE SEC/SEC OR CT448.
 GN Chlamydia trachomatis.
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
 OX NCBI_TaxID=813;
 RN [1]
 RC SEQUENCE FROM N.A.
 RP STRAIN=D/UN-3/CX;
 RX MEDLINE=9900809; PubMed=9784136;
 RA Stephens R.S., Kalman S., Lammel C.J., Pan J., Marathe R., Aravind L.,
 RA Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,
 RA Davis R.W.;
 RT "Genome sequence of an obligate intracellular pathogen of humans:
 RT Chlamydia trachomatis";
 RL Science 282:754-759 (1998).
 DR EMBL; AE001318; AAC68047.1; -;
 DR InterPro; IPR001969; Approtease site.
 DR InterPro; IPR001005; Myb DNA binding.
 DR InterPro; IPR003335; SecD_SecE.
 DR Pfam; PF02355; SecD_SecE_1.
 DR TIGRFAMs; TIGR00916; ZAO604s01; 2.
 DR TIGRFAMs; TIGR01129; secD; 1.
 DR PROSITE; PS00141; ASP_PROTEASE; UNKNOWN_1.
 DR PROSITE; PS00037; MYB_1; UNKNOWN_1.
 KW Complete proteome.
 SQ SEQUENCE 1400 AA; 156514 MW; 09400F968AD18406 CRC64;
 Alignment Scores:
 Pred. No.: 5.79e-168 Length: 1400
 Score: 2083.00 Matches: 415
 Percent Similarity: 79.28% Conservative: 67
 Best Local Similarity: 68.26% Mismatches: 80
 Query Match: 62.89% Indels: 46
 DB: 16 Gaps: 5
 US-09-868-987-1 (1-1864) x 084454 (1-1400)
 QY (2) TGGACTTCGGATATGTCAGAGGGGATCAGCGCATCTGTAATGACATATTTCTGCA 61
 Db 812 TrpThrSerLysPheCysGlnGluGlyValSerGlyThrLysAsnGlyLeuPheSerGly 831
 QY 62 AACCGTGATCGGCTAGCTGATGATGACGGTTATATGTCAGCAGCCCTATTTTA 121
 Db 832 GlyArgGlyTrpArgMetAlaValValLeuAspGlyTyrValIleSerAspProValLeu 851
 QY 122 AACGTCCTCATGAAATCATGCCAGTGTCTCAGGGAAATTTACCCCGTGAAGTGAGC 181
 Db 852 AsnValProLeuLysAspHisAlaSerValSerGlyAsnPheSerTyrArgGluValHis 871
 QY 182 AAATCGCGCTCAGATTTAAATCTGGAGCGATGCTTTTGTTCGCGAGGTCTCTCAGTGAA 241

Db 872 ArgLeuAlaAlaSerLeuLysSerGlyAlaMetSerPheIleProGluIleLeuSerGlu 891
 QY 242 GAGACGATCTCTCTGATCTTTGGGAAAAAACAATGTACACAAGGCATTATCTCAGCATGC 301
 Db 892 GluValValSerProGluLeuGlySerSerGlnArgValGlnGlyIleLeuSerValVal 911
 QY 302 TGTGGCTTGGCAATGCTTATTTTTCATGAGCGTATATTATAGATTTGGAGGCGTCATC 361
 Db 912 LeuGlyLeuValValLeuIleValLeuMetSerValTyrArgPheGlyGlyValIle 931
 QY 362 GCTTCGGAGCTGTTCTCTGAATCTTTTCTATCTGGCAGCTCTACAGTATTTCGAT 421
 Db 932 AlaSerIleAlaValLeuLeuAsnLeuLeuIleTrpAlaSerMetGlnTyrLeuAsp 951
 QY 422 GCGCACTCACCTTTGTGAGCACTCGCTGGGATTTCTTCTGATGGGATGGCGGTAGAT 481
 Db 952 AlaProLeuThrLeuSerGlyLeuAlaGlyIleIleLeuAlaMetGlyMetAlaValAsp 971
 QY 482 GCAATGTTCTTGTATTCGAAAGATCCGAGAGAAATTTTATTTGCTCTCAAGTCTTAAA 541
 Db 972 AlaAsnValLeuValPheGluArgIleArgGluGluTyrLeuLeuThrArgSerLeuSer 991
 QY 542 AAATCTGTAGAAAAGCATATACCAAGCTTTTGGAGCCATTTTGGATTTCTAAGTCTGACT 601
 Db 992 GluSerValGluAlaGlyTyrLysLysAlaPheSerAlaIlePheAspSerAsnLeuThr 1011
 QY 602 ACAGTATTTGGCTCAGCACTTTCTTCTCTAGATACAGGCGCTATTAAAGGTTTGTCT 661
 Db 1012 ThrIleLeuAlaSerAlaLeuLeuMetLeuAspThrGlyProIleLysGlyPheAla 1031
 QY 662 TTGACATGATTTTAGAAATTTTCTTCAATGTTTACGCTCTTTTCATGACTAAATTT 721
 Db 1032 LeuThrLeuIleIleGlyIlePheSerSerMetPheThrAlaLeuPheMetThrLysPhe 1051
 QY 722 TTCTTCATGCTGTGGATGAATAGACCAACATACACAGTTCGATATGATCAATAAGTTC 781
 Db 1052 PhePheValIleTrpValGlnLysThrArgLuthrGlnLeuHisMetMetAsnLysPhe 1071
 QY 782 GTGGGATAAAGCATGATTTCTTGAGAGGATGCAAAAAAATTTGGGCTGTTTCTGGAAGT 841
 Db 1072 IleGlyIleLysHisAsnPheLeuLysGluCysLysArgLeuTrpValValSerGlyVal 1091
 QY 842 GTTTTCTTTTAGGTCGCTGCTCTCGGTTTGGAGCCCTGGAATTCGTTTTCGGGAATG 901
 Db 1092 ValLeuValLeuGlyCysValGlyLeuGlyPheGlyAlaTrpAspSerValPheGlyMet 1111
 QY 902 GATTTTAAAGAGGCTATGCTTTACCTTTAATCCAAAGAGCATGCGCATCAGCGATGTT 961
 Db 1112 AspPheLysGlyGlyTyrAlaLeuThrLeuAsp 1125
 QY 962 GCT-----CAAAATCGTGGCAAGTCTTGCAATAAATCTACAGGAAGCTGGT 1006
 Db 1126 CysGluTyrAsnProGluGlnMetCysSerValLeuArgLysArgPheGlnGlnIleGly 1145
 QY 1007 CTTTCTTCTAGAGACTTCCGTTATTCAAACATTTTGGATCTTTCAGAAAAAGATCAAAATCTAT 1066
 Db 1146 LeuSerSerArgAspTyrArgValArgAlaAspSerSerGluLysValLysIleTyr 1165
 QY 1067 TTTAGTGAATAAGCTTTTAAGCTATACTAAGCAGATACGA----- 1105
 Db 1166 LeuSerGlnAsnAlaLeuAspArgValGluGlnIleGluGlyAlaLysSerGluGlnLys 1185
 QY 1106 -----GCTCTCTCTCTAAATTAACGATCATGACCTGCGGTATTGTT 1147
 Db 1186 GlySerAspTyrHisLeuAlaArgValLeuGlnValLeuSerAspSer----- 1201
 QY 1148 GGGATTGTTGTGAGAAAACAGCGCTAGATTTTCTACGGAAACTCTAAACGAAACGCAAAA 1207
 Db 1202 -----GlySerSerThrThrSerMetVal 1209
 QY 1208 TTT-----TGTCAAAGTAAGCAGCAAACTATTCGAAAGAAATTCGT 1249

Db 1210 PheAspAlaSerArgGlySerTyrPheLeuValSerGlyGlnLeuSerAsnLysMetArg 1229
Qy 1250 TATCAGCGCAGCATCGGGCTTTAGAGCTTTGGCAATCATCTGCTTANGTGAAGTTTG 1309
Db 1230 ThrGlnAlaValIleAlaLeuPheGlyAlaLeuGlyIleIleLeuPheTyrValSerLeu 1249
Qy 1310 CGCTTTGAATGGCAATATGCTTTCAGTGCATGCGATGCGCTTAAATGACCTTTGGCT 1349
Db 1250 ArgPheGlnTyrPargTyrAlaPheSerAlaIleCysAlaLeuMetHisAspLeuLeuAla 1269
Qy 1370 ACCGTGACGCTCTGTTATATGACATTTCTTTTGAAGAAATTCAAATGATTTGCA 1429
Db 1270 ThrCysAlaValIleValAlaLeuHisPhePheLeuGlnArgIleGlnIleAspLeuGln 1289
Qy 1430 GCCATGCTGCTTTATGATGATGATGGGCTTATCATTAACAATCTTGTATTTT 1489
Db 1290 AlaIleGlyAlaLeuMetThrValLeuGlyIleSerLeuLeuAsnThrLeuIleIlePhe 1309
Qy 1490 GATGCTATTCGTGAAGATCGCCAGCGAACCTGTTTACCCTATGCGATGTTTGTAT 1549
Db 1310 AspArgIleArgGlnAspArgArgGlnLysLeuPheMetProMetProIleLeuLeuAsn 1329
Qy 1550 GATCCCTCTCAAAAGACCTTCAGCGCGACGCGTAATGACAGCTACACTCTATCAGTT 1609
Db 1330 AspAlaLeuGlnLysThrLeuGlyArgThrValMetThrThrAlaThrThrLeuSerVal 1349
Qy 1610 TTGTTATGCTTTTGTATTAAGCGCGCTCTCTGCTTTAAATTTGCAATTTATGAC 1669
Db 1350 LeuValIleLeuLeuPheValGlyGlyIleSerIlePheAsnPheAlaPheIleThr 1369
Qy 1670 ATAGGAGATCTCTTACGAACTTTATCGCTCTTATATGACACACCTGTGTGTTT 1729
Db 1370 ValGlyIleLeuLeuGlyTyrLeuSerSerLeuTyrIleAlaProIleLeuLeuPhe 1389
Qy 1730 ATGCTCCGTAAAGAAATCCTCA 1753
Db 1390 MetValArgLysGlnGlnAsn 1397

RESULT 3

Q9PU4 PRELIMINARY; PRT; 1400 AA.

AC Q9PU4; 01-OCT-2000 (TrEMBLrel. 15. Created)
DT 01-OCT-2000 (TrEMBLrel. 15. Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21. Last annotation update)
DE SecDF protein, putative.
GN TC0733.
OS Chlamydia muridarum.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83560;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MOBN / NIGG;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gail S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Uterback T., Berry K., Baas S.,
RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA Gelin M., Nelson W., DeBooy R., Kolonay J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis Mohn and Chlamydia
RT pneumoniae A39";
RL Nucleic Acids Res. 28:1397-1406(2000).
DR EMBL; AE002341; AAF39543.1; -.
DR TIGR; TC0733; -.
DR InterPro; IPR001036; Acrflavin res.
DR InterPro; IPR001005; Myb DNA binding.
DR InterPro; IPR003335; SecD_SecE.
DR Pfam; PF02355; SecD_SecE; 1.
DR PRINTS; PR00702; ACRIFLAVINRP.
DR TIGRFAMs; TIGR00916; 2A0604s01; 2.
DR TIGRFAMs; TIGR01129; secD; 1.
DR PROSITE; PS00037; MYB_1; UNKNOWN_1.
KW Complete proteome.

SQ SEQUENCE 1400 AA; 156140 MW; 559E509CD823823F CRC64;
Alignment Scores:
Pred. No.: 1,08e-166 Length: 1400
Score: 2068.00 Matches: 407
Percent Similarity: 79.61% Conservative: 77
Best Local Similarity: 66.94% Mismatches: 78
Query Match: 62.44% Indels: 46
DB: 16 Gaps: 5
US-09-868-987-1 (1-1864) x Q9PU4 (1-1400)
Qy 2 TGGACCTCCGATATGTCAGAGGGATGACGGCAGCTGTAATGACAAATATTCGCA 61
Db 812 TrrPnsSerLysPheCysGlnGlnGlyValAsnGlyThrLysAsnSerGlnPheSerGly 831
Qy 62 AACCGTGATGGCGTATGCGCTAGTATGATGACGTTATATGCTACAGACCTATTTTA 121
Db 832 GlyArgGlyTrrPargMetAlaValValLeuAspGlyTyrValIleSerAspProValLeu 851
Qy 122 AACGTCCCATGAAATATCAGCCAGTGTCTCAGGAAATTTACCCACGCTGAAGTAC 181
Db 852 AsnValProLeuLysAspHisAlaSerValSerGlyAsnPheSerTyrArgGlnValAsn 871
Qy 182 AACCTGCGCTCAGATTTAAATCTGGAGCGATGCTTTGTCGCCAGGTTCTCAGTGA 241
Db 872 ArgLeuAlaSerLysPheLysSerGlyAlaMetSerPheIleProGlnIleLeuSerGln 891
Qy 242 GAGACGATCTCTTGATCTTGGGAAAAAACAATGTAACAACAAGCATTTATCAGCATGC 301
Db 892 GlnValValSerProGlnLeuGlnLysSerGlnArgLeuGlnGlyIleLeuSerValPhe 911
Qy 302 TGGGCTTGGCAATGCTTATGTTTATGATGAGCGTATATATGATTTGAGCGCTATC 361
Db 912 LeuGlyLeuValValLeuIleAlaLeuMetSerValTyrTyrLysPheGlyValIle 931
Qy 362 GCTTGGGAGCTGCTTCTGTAATCTTTGCTATGCTGAGCAGCTCAAGTATTTGAT 421
Db 932 AlaSerValAlaValLeuLeuAsnLeuLeuIleTrrPAlaSerMetGlnTyrLeuAsp 951
Qy 422 GCGCACTCACTTGTTCAGACTCGCTGGATGTTCTTGTCTATGAGGATGCGCTAGAT 481
Db 952 AlarProLeuThrLeuSerGlyLeuAlaGlyIleIleLeuAlaMetGlyMetAlaValAsp 971
Qy 482 GCAATGTTCTTGTATTCGAAAGATCCGAGAGATTTTATGTTGCTCAAGCTTTAA 541
Db 972 AlaAsnValLeuValPheGlnArgIleArgGlnGlnLysLeuLeuSerArgSerLeuSer 991
Qy 542 AAATCTGTAAGAAAGATATACCAAGGCTTTGAGCCATTTTGTATGATTCAGTACT 601
Db 992 GlnSerValGlnAlaGlyTyrLysValAlaPheSerAlaIlePheAspSerAsnLeuThr 1011
Qy 602 ACAATATGCTCAGACACTCTTCTTCTTCTAGATACAGGCGCTATTAAGGCTTTGCT 661
Db 1012 ThrIleLeuAlaSerAlaLeuLeuLeuMetLeuAspThrGlyProIleLysGlnPheAla 1031
Qy 662 TTGACATGATTTTAGAATTTTCTCTTCATGTTTACGCTCTTTTCATGACTTAAATTT 721
Db 1032 LeuThrLeuIleIleGlyIlePheSerSerMetPheThrSerLeuPheMetThrLysPhe 1051
Qy 722 TTCTCATCGTGGATGATTAAGCCACCATACACAGTGTGATGATGATTAAGTGA 781
Db 1052 PhePheLeuIleTrrPleGlnLysThrArgGlnThrGlnLeuHisPheMetAsnLysPhe 1071
Qy 782 GTGGGGATTAAGCATGATTTCTTGAGAGATGCAAAAACCTTTGGCGCTGTTCTGGAAGT 841
Db 1072 IleGlyValLysHisAsnPheLeuLysGlnCysLysArgLeuTrrPValValSerGlyAla 1091
Qy 842 GTTTTCTTTTAGGTTGCTGCTCTCGGGTTTGGAGCTCGAATTCGCTTTTGGGAATG 901
Db 1092 ValValValLeuGlnCysIleGlyLeuGlyPheGlyValAlaTrrPAspSerIleLeuGlyMet 1111
Qy 902 GATTTTAAAGAGGATATGCTTACCTTTAATCCAAAAGACATGGCAGCATGTT 961

QY	424	GCACCTACCCCTTGGCAGACCTGGCTGGGATGTTCTTTCGATACAGGGATGGCCGTGATGAC	483
Db	439	rTthleuthrleuthrproctylleaaaglyllevalleuthrilleeglymeAlaValaAspse	459
QY	484	AAATGTTCTTGATTCGAAAATAATCCGAGAGAAATTTATGTCTCAAAAGCTTTAAAAA	543
Db	459	rAsnValleuLeuLeuYrGluTrgIleaaArgIuGluValLysseArgLysProLeuIleGI	479
QY	544	ATCTGTAGAAAAAGGATATTCACAAAGCTTTGGAGCCATTTTGGATCTTACCTGATACAC	603
Db	479	nAlaLeuAspAsnGlyPheThrAlaGalaPheAlaThrIlelleAspAlaAsnLeuThrTh	499
QY	604	AGTATGGCCCTCACACACTCTCTTCTTCCTCAGATATACAGGGCCCTATTTAAAGGGTTGCTT	663
Db	499	rLeuIleValAlaAserValleuPheUryMetGlyThnGlyProValLysGlyPheAlaVa	519
QY	664	GACATTTGATTTTACGAAATTTCTCTTCATAGTTTACGGCTCTTTTCATGACATAATTTT	723
Db	519	lThrIleuAlaValGlyIleThrThrValPheThrAlaYrThleuThrAlaTrpMe	539
QY	724	CTTCATGCTGTGATGATGATGATAGAACCCAACTACACAGTTGGCAT-----ATGATGAA	774
Db	539	cPheGlyYrTrpValAlaArgAserAlaArgProLysHleuProLysGlyValaArgThrAl	559
QY	775	TAAGTTCGTGGGATPAAAGCATGATTTCTTGAGAGATGCACAAAACTTTGGCGTGTTC	834
Db	559	aMetPheAspIlyrArgAspIleProPheMetCArgYrArgArgValaValPheMetIleTh	579
QY	835	TGGAAGCTTTTCTTTTATGATGGTGGCTGCTCGGGTTTGGAGCGCTGGAAATTCGGTTT	894
Db	579	rGly---ValIleLeuLeuLacysValGly---GlyPheValAlaLysGlyLeuAsnIle	597
QY	895	GGGAATGATTTTAAAGAGAGGTATGCTTACTTAACTTCAAAAGACATGGCATCAG	954
Db	597	uGlyIleAspPheGlnGlySerValIleGluValaArgAlaGln---GlyGluAl	616
QY	955	CGATGTTGCTCAAAATGCGTGGCAAAAGTTTGACATAAACTACAGSAAAGCTGGTCTTCTTC	1014
Db	616	aAspLeuAlaAspIleArg-----GluArgLeuAsnGlnLeuAsnLeuGly--	631
QY	1015	TAGAGACTTCGGTATCAAAACATTTGGATCTTCAGAAAAGATCAAAATATATTTTAGTA	1074
Db	632	---GluIleGlnAlaGlnAsnPheGlySerProGlnAspValleuIleArgIle----	648
QY	1075	TAAAGCTTTAAGCTATACATACAGCATAGACGCTCTCTCTTAAATTAACATCATGAG	1134
Db	649	-GlnAlaGlnAspIlyGlyGlnAsnAlaGlnGlnSerAlaIleThrLeu-----	664
QY	1135	CTGGCGTTATGTGGGATTTGTTCAGAAAACAGCCCTAAGTTTCTTCAGSAAACTTAA	1194
Db	665	-----ValaArgIlyGluLeu-----GluAspIy	672
QY	1195	ACGAAACGCAAAATTTTGTTCAAAGGTATACAGACAAACTATCGAAGAAAATGCGTTATCA	1254
Db	672	sTyraAspPheArgIyAlaGluValValGlyProAlaValaIserGlyAspLeuThrValTrh	692
QY	1255	GGCGACCATCGGGCTTTTAAAGCTTTGGCAATCATCTTGCTATAGTGAAGTTGGCGTT	1314
Db	692	rSerThrIleGlyValleuLeuAlaMetAlaIleIleMetValTyrIleTrpValaArgPh	712
QY	1315	TGAATGGCAATATGCTTCAAGTGGCGGATCGCTTAAATTCATGACCTTTTGGCTAACCG	1374
Db	712	eGluTrpGlnPheAlaLeuAlaValaValIleSerMetValHisAspValaValaPheThr--	731
QY	1375	TGCAGTCTTGTTPAGACATTTCTTTTGAAGAAAATTCAAATAGATTTGCAAGCAT	1433
Db	732	-----IleGlyLeuPheValPheLeuGlyIleGluPheAsnLeuThrSerIl	747
QY	1435	TGGGCTTATATGACTGATATGGGGGTATTCATTAACAATCACTTGATCATTTTGGATG	1494
Db	747	eAlaIleIleuThrIleIleGlyIySerLeuAsnAspThrValaValIleTrpAspTr	767

```
Db 317 GlnIleArgGluProIleLeuGlyGlyThrGlyGlnIleSerGlyAsnPhetThrAlaGln 336
Qy 173 GAAGTGACAAACCTCGCTCGAGATTAAATCTGGAGCGATCTCTTTGTTCCCGAGGTT 232
Db 337 SerAlaAsnAspLeuAlaValLeuLeuArgAlaGlyAlaLeuProAlaThrLeuThrVal 356
Qy 233 CTCAGTGAAGAGACGATCTCTCTGATCTTGGGAAAAAACAAATGACACAGGCATTATC 292
Db 357 IleGluGluArgThrValGlyProGlyLeuGlyGlnAspSerIleHisAlaGlyLysVal 376
Qy 293 TCAGCATGCTGGCTGGCAATGCTTATTGTTGATCAGCGGTATATATAGATTGGA 352
Db 377 AlaGlyIleIleGlySerIleLeuValAlaAlaPheMetPheValAlaThrGlyPheLeu 396
Qy 353 GCGCTCATCGCTTCGGGAGTGTCTCTTGAATCTTTGCTATCTGGCGAGCTCTACAG 412
Db 397 GlyPheLeuAlaAsnIleAlaLeuAlaValHisValAlaMetIleValGlyLeuLeuSer 416
Qy 413 TATTGGATGGCCCACTCACCTTGTGACAGCTCGCTGGGATGTTCTTGTGATGGGATG 472
Db 417 LeuLeuGlyAlaThrLeuThrLeuProGlyIleAlaGlyIleValLeuThrIleGlyMet 436
Qy 473 GCGGTAGATGCAATGTTCTGTATTCGAAAGAACCCGAGAGGATTTTATTGTCTCAA 532
Db 437 AlaValAspSerAsnValIleThrGluArgIleArgGluGluArgAlaGlyArg 456
Qy 533 AGTCTTAAAAATCTGTAGAAAAAGATATACCAAGGCTTTTGGAGCCATTTTGTATCT 592
Db 457 SerValIleGlnAlaIleAspThrGlyPheSerLysAlaLeuAlaThrIleValAspSer 476
Qy 593 AACTTGACTACAGTATGGCTCAGCACTCTTCTTCTTCTAGATACAGGCGCTATTAAA 652
Db 477 AsnValThrSerLeuIleAlaThrValValLeuPheThrLeuGlyThrGlyProValLys 496
Qy 653 GGGTTGCTTGTGACATGATTTAGGAATTTCTCTTCAATGTTTACGGCTCTTTTCATG 712
Db 497 GlyPheAlaIleThrTyraIleGlyIleLeuThrValPheThrAlaPheThrPhe 516
Qy 713 ACTAAATTTTCTTCATGCTGGATGAATTAAGACCAACATACACAGTGTGCATATGATG 772
Db 517 ThrArgLeuLeuValSerIleThrLeuArgArgAlaArgProLysGluLeuProLysAla 536
Qy 773 AAT---AAGTTCGTG-----GGGATAAGCATGATTTCTTGAGAGGATGCAAAAAAATT 823
Db 537 ProValThrPheIleProGlyThrLysIleProPheMet---GlyIleArgArg--- 554
Qy 824 TGGGCTGTTCTGGAACTGTTTCTTTTAGTTCGGTTCGCTCGGTTTGGAGCCCTGG 883
Db 555 TrpThrPheAlaLeuSerSerPheLeuSerValLeuSerValValLeuPheMetThrVal 574
Qy 884 AATTCGCTTTTGGGAATGGATTTTAAAGAGGATGATGCTTTTACCTTTAATCCAAAAGAG 943
Db 575 AspIleAsnThrGlyIleAspPheLysGlyLysMetIleGluValLysSerLysSer 594
Qy 944 CATGGCATCAGCATGTTGCTCAATGCGTGGCAAGTTGTGCATAAACTACAGGAAGCT 1003
Db 595 ---GlyAspAlaAsnLeuGlyAspIleArgAlaArgLeuMetGlu----- 608
Qy 1004 GGTCTTCTTCTAGAGATTCGGTATTCAACATTTGATCTTTCAGAAAAGATCAAAATC 1063
Db 609 ---LeuAsnIleGlyGluValGlnValGlnGlnPheGlyAlaPro----- 622
Qy 1064 TATTTTAGTATAAGCTTTAAGCTATATAAGCAGATACAGGCTCTCTCTCTAAATA 1123
Db 623 -----AsnAspValLeuIleArgValGlyThrGlnAspGlyGlyGluAsnAlaGln 640
Qy 1124 ACGATCATGAGCTGGCGTTATTGTGGGATTTGTGTGAGAAAACAGGCTAGATTTCTCTAC 1183
Db 641 ThrValIleAsp-----LysValArgGlyGluLeuGlnAspGlnTyr 654
Qy 1184 GGAAACTCTAAACGAAACGAAAATTTTGGTCAAGAGTAAGCAGCAACTATCGAAGAAA 1243
::: ::::
```

```
Db 655 ---AspPheArgArgValGluValValGlyProThrValSerGlyGluLeuAlaLys--- 672
Qy 1244 ATGCGTTATCAGCGGACCATCGGCTTTTAGAGCTTTGGCAATCATCTTCTCTATGTG 1303
Db 673 -----GlnGlyThrIleAlaMetLeuValAlaLeuValGlyIleLeuValThrVal 689
Qy 1304 AGTTTGGCTTTGAATGGCAATATGCTTTCAGTCCGTATCGCTTTTAATTCATGACCTT 1363
Db 690 TrpPheArgPheGluThrPgnPheAlaValGlyAlaIleValAlaThrValHisAspVal 709
Qy 1364 TTGCGTACCTGTCGAGCTGTTTATAGCACATTTCTTTTGAAGAAAATTCAAATAGAT 1423
Db 710 ValMetThr-----LeuGlyPheValIleThrGlyLeuGluPheAsn 724
Qy 1424 TTGCAAGCCATTTGCTGCTTTAATGACTGTATTGGGTATTTCATTAACAATCTTTGATC 1483
Db 725 GlnSerSerLeuAlaAlaIleLeuThrIleIleGlyThrSerLeuAsnAspThrIleVal 744
Qy 1484 ATTTTGTATGCTATTGCTGAAGATCGCAAGCGAACCTGTTTACCCCTATGTCATGTTT 1543
Db 745 ValTyraAspArgValArgGluAspLeuArgLysTyraLysArgMetProLeuProGlnLeu 764
Qy 1544 GTTAATGATGCCCTTCAAAGACGCTTCCAGCGCAGCGTAATGACAAACAGCTTCAACTCTA 1603
Db 765 LeuAsnAsnAlaIleAsnGluThrLeuSerArgThrThrLeuThrSerValThrThrIle 784
Qy 1604 TCAGTTTGTGTTGCTTTTATAGCGGCTCTCTGCTTTTAAATTTTGCATTATT 1663
Db 785 LeuAlaLeuLeuAlaLeuValLeuPheGlyGlyGluValIleArgSerPheThrMetAla 804
Qy 1664 ATGACCATAGGATTTCTTAGGAATTTTATGCTCTCTTTATATATGACCACTCTGTTG 1723
Db 805 MetLeupheGlyValValPheGlyThrTyraSerIlePheIleAlaAlaProLeuLeu 824
Qy 1724 ---TTGTTTATGCTCGCTAAAGAA 1744
Db 825 IleLeupheLysLeuArgProGln 832
RESULT 6
Q8YHM6 PRELIMINARY; PRT; 845 AA.
ID Q8YHM6
AC Q8YHM6;
DT 01-MAR-2002 (TEMBLrel. 20, Created)
DT 01-MAR-2002 (TEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TEMBLrel. 21, Last annotation update)
DE Protein translocase subunit SECD / protein translocase subunit
DE SCF.
GN BMEI0680.
OS Brucella melitensis.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Brucellaceae; Brucella.
OX NCBI_TaxID=29459;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=16M / ATCC 23456 / BIOTYPE 1;
RX MEDLINE=20020109; PubMed=11756688;
RA DelVecchio V.G., Kapural V., Redkar R.J., Patra G., Mujar C., Los T.,
RA Ivanova N., Anderson I., Bhattacharya A., Lykidis A., Reznik G.,
RA Jablonski L., Larsen N., D'Souza M., Bernal A., Mazur M., Goltzman E.,
RA Selkov E., Elzer P.H., Hagius S., O'Callaghan D., Letesson J.-J.,
RA Haselkorn R., Kyrides N., Overbeek R.;
RT "The genome sequence of the facultative intracellular pathogen
RL Brucella melitensis."
RL Proc. Natl. Acad. Sci. U.S.A. 99:443-448 (2002).
DR EMBL; AE009509; ALU51861.1; -.
DR InterPro; IPR003335; SecD_SecF.
DR Pfam; PF02355; SecD_SecF; 1.
DR TIGRFAMs; TIGR00916; 2A0604e01; 2.
DR TIGRFAMs; TIGR00965; 3a0501e07; 1.
DR TIGRFAMs; TIGR01129; secD; 1.
KW Complete proteome.
SQ SEQUENCE 845 AA; 9208 MW; 1397716197787368 CRC64;
```

Alignment Scores:

Pred. No.:	4,086-52	Length:	845
Score:	715.50	Matches:	175
Percent Similarity:	52.87%	Conservative:	129
Best Local Similarity:	30.43%	Mismatches:	198
Query Match:	21.60%	Indels:	73
	16	Gaps:	14

US-09-868-987-1 (1-1864) x Q8YH6 (1-845)

```

QY 80 GCTGTAGTATGACGGTTATATGTCAGACCCCTATTATTAAGCTCCATGAAA--- 136
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 298 AlalleValleuaspGlyValleuSerAlaProValIleArgGluProIleThrGly 317
QY 137 AATATGCCAGTGTCTCAGGAAATTAACCAAGCTGATGAGCAAACTGCCTCAGAT 196
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 318 GlySerGlyGlnIleSerGlySerPheThrValGlnAspThrValIleuSerAlaLeu 337
QY 197 TTAATATCTGAGGATGTCCTTTGTTCCGAGGTCCTCATGGAAGAGAGATCTCTCT 256
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 338 LeuArgAlaGlyAlaLeuProAlaProLeuThrValIleGluGluArgThrValGlyPro 357
QY 257 GATCTGGGAAATAAACAATGTAACAAGCATTAATCTCAGCATGCTGTGCTGGCAATG 316
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 358 AspleuGlyGlyAspAlaIleLysMetGlyLeuMetThrGlyIleIleGlyPheLeuLeu 377
QY 317 CTTATGTTTGGATGAGCGTATATTAAGATTTGGAGCGCTCATGCTTCGGAGCTGTT 376
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 378 ValAlaValPheIleLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 397
QY 377 CTTCTGATCTTTGCTTATCTATGCGGAGCTTACAGTATTTGGATGGCCAGCTCACTG 436
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 398 LeuLeuHisThrLeuLeuThrPheSerAlaLeuSerLeuIleGlyAlaThrLeuThrLeu 417
QY 437 TCAGAGCTCGCTGGATGTTCTTCTGATGGGATGGCCGTAGATGCAATGTTCTGTA 496
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 418 ProGlyIleAlaGlyIleIleLeuGlyIleGlyIleAlaValAlaAsnIleLeuIle 437
QY 497 TTCGAAAGATCCGAGGAAATTTTATGTCCTCAAGCTTTAAATAATCTGTGAAAAA 556
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 438 AsnGluArgIleArgGluGluThrArgLysGlyLeuGlyAlaMetAlaIleuAspLys 457
QY 557 GGATATACCAAGCTTTTGGAGCATTTTGTATCTAATCTGATCAGTATGGCCCA 616
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 458 GlyPheHisSerAlaPheAlaThrIleValAspAlaAsnValThrThrLeuThrAlaThr 477
QY 617 GCATCTTCTTCTCTCTATACAGAGGCTTAAAGGTTGCTTGACATTTGATTTTA 676
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 478 IleLeuLeuPheLeuPheGlyThrGlyProValArgGlyPheAlaValThrMetMetLeu 497
QY 677 GGAATTTCTCTCAATGTTTACGGCTCTTTTCATGACTAAATTTTCTTCATGCTGTG 736
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 498 GlyIleAlaIleSerMetPheThrAspValThrLeuValArg---MetIleMetAlaTrp 516
QY 737 ATGAATAG-----ACC 748
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 517 PheValAlaArgArgArgLysLeuLysValLeuHisIleGluProPheLeuLysPheValPro 536
QY 749 CAATATACAGATGTCATATGATGAT---AAGTTCGTGGGATTAAGCATGATTTCTTG 805
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 537 GluHisThrAsnPheArgPheMetAsnAlaArgPheIleGlyIle----- 551
QY 806 AGAGGATGCAAAAACTTT36GCTGTTTCTGGAAGTGTGTTTCTTTAGGTGCGTGTCT 865
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 552 ---GlyValSerIleValLeuSerIleAla---SerIleIleLeuPhePheLysProGly 569
QY 866 CTCGGGTTTGGAGCTGGAATTCGCTTTTGGGAATGATTTTAAAGAGGGTATGCTTCT 925
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 570 LeuAsnLys-----GlyIleAspPheLysGlyGlyIleGlnAla 582
QY 926 ACCTTAATCCAAAAGAGCATGATCAGCATGTTGCTCAAAAGCGGCAAGATTTGTG 985
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 583 GluIleSerThrSerGlnPro-----AlaAspLeuAlaGlnIleuArgAlaLysLeu--- 599

```

```

QY 986 CATTAATAAGAGAGAGCTGCTTTTCTCTAGAGACTTCGATATTCAAACTTTGATCT 1045
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 600 -----GlyAlaLeuAsnLeuGlyValAlaAlaLeuGlnThrAlaGlySer 614
QY 1046 TCAGAAAGATCAAAATCTATTTT-----AGTATTAAGCTTTAAGC 1087
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 615 ProAsnGlnValLeuIleArgValGlnArgGlnGlyGlyGluGluAlaGlnThrAla 634
QY 1088 TATATCAAGCAATACAGAGCTCTCTCTAAATTAACGATCATGAGCTGACGTTATG 1147
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 635 AlaIleAsnLysMetArgGluAlaValThrGluLeuAsp-----Pro 648
QY 1148 GGGATTTGCTCAAAAACAGAGCTTATTTCTTACGAAACTCTAAAGCAAGCAAAA 1207
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 649 GlyValIle-----GluArgThrGluVal 657
QY 1208 TTTTGGTCAAGGTAAGCAAGCAACTATCGAAGAAATGCGTTATCAGGCACTCGG 1267
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 658 ValGlyProLysValSerGlyGluLeuAlaArg-----SerGlyMetIleAla 673
QY 1268 CTTTAAAGACCTTTGGCAATCATCTGCTCATGAGTTGGCTTTGAATGCAATAT 1327
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 674 ValIleLeuSerAlaValAlaMetLeuPheTyrlleTrpArgPheGluTrpPhe 693
QY 1328 GCTTTCAGTCCGCTATGCGCTTTAATTCATGACCTTTTGCTACCTGTGACGCTGTT 1387
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 694 AlaLeuGlyAlaIleAlaThrLeuIleLeuAsp-----ThrThrLysIle 708
QY 1388 ATAGCAATTTCTTTTGAAGAAATTCAAATGATTTGCAAGCCATGGCTTTATATG 1447
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 709 IleGlyPhePheAlaLeuThrGlnLeuAspPheAsnLeuThrAlaIleAlaLeuLeu 728
QY 1448 ACTGATTTGGGATTTTCATTAACAATCACTTGATCATATTTGATGCTATGCTGAAGAT 1507
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 729 ThrIleIleGlySerValAlaAsnAspLysValValIleTyrsArgPheArgGluAsn 748
QY 1508 CGCCAGCGCAAGCTGTTTACCCCTATGATGCTTTTATGATATGATCCCTTCAAAAGAGC 1567
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 749 MetArgLeuTyrlsSerIleThrLeuArgGluIleIleAspMetSerIleAsnGlnVal 768
QY 1568 TTCAGCGGCAAGGTAATGACACAGCTACACTCATGATTTGTTGTAATGCTTTGTT 1627
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 769 LeuValArgGlyValIleTyrlsThrSerMetThrThrPheLeuLysMetPheProMetAlaIle 788
QY 1628 ATAGGCGCTCTCTCTCTTAAATTTGCAATTTATTAAGCAATGAGATTTCTTAGAGA 1687
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 789 TrpGlyGlySerAlaValHisAsnPheAlaValProMetLeuPheGlyIleValIleAla 808
QY 1688 ACTTATTCGCTCTTATATGACACCACTCTGTGTTGTTATG 1732
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 809 ThrSerSerSerIlePheIleAlaIleAlaProIleLeuLeuLeuLeu 823

```

RESULT 7

099TL6

ID 099TL6 PRELIMINARY; PRT; 759 AA.

```

AC 099TL6;
DT 01-JUN-2001 (TREMblrel. 17, Created)
DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)
DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
DE Protein-export membrane protein SecDF.
GN SECDF OR SAVA1637 OR SAI1463.
OS Staphylococcus aureus (strain Mu50 / ATCC 700699), and
OC Staphylococcus aureus (strain N315).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Staphylococcus.
OX NCBI_TaxID=158878, 158879;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=S.aureus (strain Mu50), and S.aureus (strain N315);
RX MEDLINE=21311952; PubMed=1148146;
RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
Cui L., Oguchi A., Aoki K.-I., Nagai Y., Lian J.-Q., Ito T.,

```


Qy	299	TGCTGTGGCTTGGCAATGCTTATTGTTTGTATGACGGCTATTATTATAGATTTTGGAGCGCTC	358
Db	371	LeuAlaGlyPheGlyLeuValValLeuMetValValLeuTyrGlyAlaTrpGlyMet	390
Qy	359	ATCGCTTCGGAGCTGTTCTTCTGAATCTTTTGGCTTATCTGGCAGCTCTACAGATTG	418
Db	391	IleAlaAsnValGlyLeuValLeuHisThrIleMetThrIleGlyValLeuGlyLeuIle	410
Qy	419	GATGCGCCACTCACCTTGTTCAGACTCGCTGGGATGTCTTGTCTATGGGATGGCGCTA	478
Db	411	GlySerThrLeuThrLeuProGlyIleAlaGlyIleIleLeuGlyIleGlyMetAlaVal	430
Qy	479	GATGCAAAATGTTCTGTATTCGAAAGAAATCCGAGAGGAATTTTATTGTCTCAAAAGTCTT	538
Db	431	AspAlaAsnIleLeuIleAsnAlaArgIleAsgGluThrGluAlaGlyAlaGlyAla	450
Qy	539	AAAAATCTAGAAAAAGGATATACCAAGGCTTTTGGAGCCATTTTGTATTCTTAACCTTG	598
Db	451	MetLysAlaLeuAspIleGlyPheAsnLysAlaThrIleIleAspSerAsnVal	470
Qy	599	ACTACAGTATTGGCCCTCAGCACTCTCTTTCTTCCTAGATACAGGCGCTATTAAAGGGTTT	658
Db	471	ThrThrLeuAlaGlyThrValLeuLeuPheTrpPheGlySerGlyProValLysGlyPhe	490
Qy	659	GCTTTGACATTGATTTTAGGAATTTTCTCTCAATGTTTAGCGCTCTTTTCATGACTAAA	718
Db	491	AlaValThrMetMetLeuGlyIleGlyIleSerMetPheThrSerValThrValValArg	510
Qy	719	TTTTTCTTCATGCTGTGGATGAATAGACCCAAACATACACAGTTGCATATGATGAATAAG	778
Db	511	LeuLeuMetArgLysValValValArgLysMetLysLeuGluIleProSerLeu	530
Qy	779	TTCTGTTGGGATAAAGCAT-----GATTCTTTGAGAGATGCAAAAACCTTTGG	826
Db	531	PheLysValProGlnLeuProThrPheSerPheMetLysGly-----	545
Qy	827	GCTGTTTCTGGAAGTGTTTTCTTTAGGTTTCGGTGTCTTCGGGTTTGGAGCTCGGAAT	886
Db	546	-----ArgPheLeuAlaIleGlyPheSerAlaPheLeu	556
Qy	887	TCCGTT-----TTGGGAATGGATTTTAAA	910
Db	557	SerLysSerValIleLeuPheThrProGlyLeuAsnTyrGlyIleAspPheIle	576
Qy	911	GGAGGTATGCTTTACCTTTAATCCAAAGAGCATGCGATCAGCGATGTGCTCAATAG	970
Db	577	GlyGlyIleGlnValGluAlaValSerLysGluLys-----IleAsnLeuProThrLeu	594
Qy	971	CGTGGCAAACTGTGCTATAACTACAGGAAGCTGCTCTTCTTCTAGAGACTCCCGTATT	1030
Db	595	Arg-----GlnSerLeuGluLeuAsnLeuGly-----GluValAlaLeu	608
Qy	1031	CAAACTTTGATCTTCAGAAAAGATCAAAATCTATTATAGTGATAAA-----	1078
Db	609	GlnAspPheGlyGlyGlyGlnSerValLeuIleAsgValGlnAsgGlnProGlyGlyGlu	628
Qy	1079	-----GCTTTAGCTATACTAGACAGATACAGAGCTCTCTCCTAAATTA	1123
Db	629	GlnAlaGlnThrValAlaLeuAsnLysValLysAspAlaValThrThrAlaIleProGly	648
Qy	1124	ACGATCATGAGCTGGCGTTATTGTGGATTGTGTGCAGAAACAGCGCTTAGATTTCTCTAC	1183
Db	649	AlaSerMet-----	651
Qy	1184	GGAACTCTAAACGAAACCAAAATTTTGGTCAAAAGGTAAGCAGCAAACTATCGAAGAAA	1243
Db	652	-----GluArgThrGluValValGlyProThrValSerGlyGlyLeuAlaAsnSer	668
Qy	1244	ATGCGTTATCGCGGACCATTCGGCTTTTAGA-----GCTTTGGCAATCATCTTG	1294
Db	669	-----GlyPheLeuAlaValAlaLeuAlaMetLeuAlaIleLeu	681

QY 1295 CTCTATGAGATTGCGCTTGAATGCAATATGCTTCAGTCCGATGAGCTTATTT 1354
 |||||
 DB 682 LeuYr11etpPheArxPheC1urpHisPheAlaValGlyAlaIleAlaValIleu 701
 QY 1355 CATGACCTTTGGGACTGTCGACCTTTGTTATAGACATTTCTTTGGAAGAAATT 1414
 |||||
 DB 702 LeuAsp11etpHrLysThr-----ValGlyPhePheAlaLeuThrGlyLe 716
 QY 1415 CAATATGATTTGCAAGCCATGTCGCTTATATGACTGATGAGGATTCATTAACAT 1474
 |||||
 DB 717 AspPheAlaLeuThrAlaIleAlaIleAlaLeuThrMetIleGlyTyrSerValAsnAsp 736
 QY 1475 ACTTGTATCATTTTGTATGATTTGATTTGATTAAGATCCCAAGCAACCTGTTAACCCCTAAG 1534
 |||||
 DB 737 LysValValValValTyrAspArgMetArgGluAsnLeuArgLysTyrLysSerMetProhe 756
 QY 1535 CATGTTATGATATGATGAGCCCTTCAAAAGACGTTACAGCCGACGGATATGCAACAGCT 1594
 |||||
 DB 757 SerAspLeuIleAspMetSerIleAsnGlnValIleAlaArgCysIlePheThrSerAla 776
 QY 1595 ACAACTATATGATGATTTGCTCAATGCTTTGTTATAGCGCGCTCCTGCTTTATATTTT 1654
 |||||
 DB 777 AlaThrAlaLeuSerLeuValProMetAlaIleTfpgLysGlyAlaValAlaGlnSerPhe 796
 QY 1655 GCATTATATGACCATGAGGATTTCTTATAGCACTTATGCTCTTTATATATGACACA 1714
 |||||
 DB 797 AlaTyrProMetIlePheGlyValIleValAlaThrThrSerSerIleTyrIleGly 816
 QY 1715 CCTCGTGTGTTGTTATGCTCCGT 1738
 |||||
 DB 817 ProIleLeuLeuPheLeuSerArg 824

RESULT 10

032047 PRELIMINARY; PRT; 737 AA.

AC 032047; (TREMBLER, 05, Created)

DT 01-AUG-1999 (TREMBLER, 11, Last sequence update)

DT 01-JUN-2002 (TREMBLER, 21, Last annotation update)

DE SECDF protein (Protein-export membrane protein).

GN SECDF OR SECDF.

OS Bacillus subtilis.

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;

OC Bacillaceae; Bacillus.

OX NCBI_TaxID=1423;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=168;

RX MEDLINE=98362000; PubMed=9694879;

RA Bolhuis A., Broekhuizen C.P., Sorokin A., van Roosmalen M.L.,

RA Verema G., Bron S., Quak W.J., Van Dijk J.M.,

RT "SecDF of Bacillus subtilis, a molecular Siamese twin required for the

RT efficient secretion of proteins";

RL J. Biol. Chem. 273:21217-21224(1998).

RN [2]

RP SEQUENCE OF 167-737 FROM N.A.

RC STRAIN=168;

RX MEDLINE=98044033; PubMed=9384377;

RA Kunst F., Ogasawara N., Moser I., Albertini A.M., Alloni G.,

RA Azevedo V., Bertero M.G., Besieres P., Bolotin A., Borchert S.,

RA Brouillet R., Boursier L., Brans A., Braun M., Bignell S.C., Bron S.,

RA Choi S.K., Codani J.J., Conerton I.F., Cummings N.J., Daniel R.A.,

RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,

RA Entian K.D., Errington J., Fabret C., Ferrari B., Foulger D.,

RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,

RA Gilm S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,

RA Giuseppe G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,

RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,

RA Joris B., Karamata D., Kasahara Y., Klaerr-Blandhard M., Klein C.,

RA Kobayashi Y., Koester P., Koningsstein G., Krogh S., Kumano M.,

RA Kunita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,

RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,

RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
 RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Pak S.H.,
 RA Paro V., Pohl T.M., Portetelle D., Potwolk G., Prescott A.M.,
 RA Pressac E., Pujic P., Purnelle B., Rapoport S., Rey M., Reynolds S.,
 RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadate Y.,
 RA Sato T., Scanlan E., Schleich S., Schreier R., Scoffone F.,
 RA Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,
 RA Sorokin A., Taccioni E., Takagi T., Takahashi H., Takemaru K.,
 RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
 RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
 RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenger T.,
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
 RA Yoshida K., Yoshikawa H.F., Zimetstein E., Yoshikawa H., Danchin A.,
 RA "The complete genome sequence of the gram-positive bacterium Bacillus
 RT subtilis";
 RL Nature 390:249-256(1997).
 RN [3]
 RP SEQUENCE OF 167-737 FROM N.A.
 RC STRAIN=168;
 RA Kunst F., Ogasawara N., Yoshikawa H., Danchin A.,
 RA Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF024506; AAC31122.1; -
 DR EMBL; Z99118; CAB14724.1; -
 DR InterPro: IPR001036; SecDF_vin res.
 DR InterPro: IPR003335; SecDF_SecF.
 DR Pfam; PF02355; SecDF_SecF; 1.
 DR PRINTS; PR00702; ACRIPLAVINRP.
 DR TIGRFAMs; TIGR00916; 2A0604801; 2.
 DR TIGRFAMs; TIGR01129; secD; 1.
 KM Complete proteome.
 SQ SEQUENCE 737 AA; 81653 MW; 54C397E93979177 CRC64;

Alignment Scores:

Pred. No.: 5.5e-49 Length: 737
 Score: 678.50 Matches: 169
 Percent Similarity: 50.43% Conservative: 125
 Best Local Similarity: 28.99% Mismatches: 192
 Query Match: 20.49% Indels: 97
 DB: 16 Gaps: 15

US-09-868-987-1 (1-1864) x 032047 (1-737)

QY 104 GTCCAGAGCCCTATTAAAGTCCATGAAA--AATCATGCCAGTGTCTGAGGAAA 160
 |||||
 DB 197 ValSerAlaProAsnValSerGlnGlnLeuAsnThrAspValLysIleGlnGlnHis 216
 QY 161 TTATACCAACCGTGAAGTGAGCAAACTCGCCTCATGATTTAAATCGAGGAGTCTTTT 220
 |||||
 DB 217 PheThrAlaGlnGlnValLysAspLeuAlaSerIleLeuAsnAlaGlyAlaLeu----- 234
 QY 221 GTTCCGAGGTTCTCACTGAAGAG-----ACGATCTCTTGTGATCTTGGAAAAA 271
 |||||
 DB 235 ---ProValLysLeuThrGlnLysTyrSerThrSerValGlyAlaGlnPheGlnGln 253
 QY 272 CAATGTACACAGGATTAATCTCAGCATGCTGTGGCTTGCAAGCTTATGTTTATG 331
 |||||
 DB 254 AlaLeuHisAspThrValPheAlaGlyIleValGlyIleAlaIleIlePheLeuPheMet 273
 QY 332 AGCTATATATATGATTGGAGGCGTCATCGCTTCGGAGCGTGTCTTGAATCTTTTG 391
 |||||
 DB 274 LeuPheTyrTyrArgLeuProGlyLeuIle-----AlaValIleThrLeuSerVal 290
 QY 392 CTATATGCGGAGCTTACAG-----TATTGATGCGGACATCACCTTGTACAGA 442
 |||||
 DB 291 TyrIleTyrIleThrLeuGlnIlePheAspTyrPheAsnAlaValLeuThrLeuProGly 310
 QY 443 CTCCTGCGGATTTCTTCTGATAGGAGTGGCCCTAGATGCAATGTTCTTGTATGAA 502
 |||||
 DB 311 IleAlaAlaLeuIleLeuGlyValGlyMetAlaValAlaAspAlaAsnIleIleThrGln 330
 QY 503 AGATTCGAGAGGAGATTTTATGTTCTCAAGTGTAAAGTGTAAATCTGTAGAAAAGATAT 562
 |||||
 DB 331 ArgIleLysGlnGlnLeuLysLeuGlyLysSerValArgSerAlaPheArgSerGlyAsn 350


```

|||||
Db 270 LeuysserglyalaleuProilleysilevalThralaaleuargserilleglyPro 289
QY 257 GATCTGGGAAAAAACAATGACACAGCATTTATCTGAGCATGCTGGCTGGCAATG 316
Db 290 SerleuglylaasparaileargserglyalaleuValleuValleuValleu 309
QY 317 CTATTGTTTGAAGGCGTATATAT---AGATTGAGGCGTCATGCTTCGGAGCT 373
Db 310 ValphevalmetleuphealatyryglyleuptrpheglyleuValglyAlaleu 329
QY 374 GTTCTTCGAAATCTTTGCTTGGGAGCTCTACAGTATTGGATGGCCACTCACC 433
Db 330 LeuuepneserillelleuuglylleuuglyglypheglyAlathleuThr 349
QY 434 TTGTACAGACTCGTGGGATTGTTCTGTATGGGATGGCCGATGACAAATGTTCT 493
Db 350 LeuproglyllealaglyleuValleuThrilleglyAlaValaspglyAsnValle 369
QY 494 GTATTGCAAGATCCGAGAGATTTTATTTCTCAAGCTTTAAAAATCTGTAGA 553
Db 370 SerphegluargyleglyleuAlaargglylysglylleysAsnAlalle 389
QY 554 AAAGGATATACCAAGGCTTTGGAGCATTTTGTATCTCACTTACATGATATGGCC 613
Db 390 AlaelytyrGluHsSerThralaAlalleuAspValasnaHsleuSer 409
QY 614 TCAGCACTCTTTCTCTCTAGATACAGGCGCTATTAAGGGTTTGTGACATGAT 673
Db 410 AlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 429
QY 674 TTAGGAATTTCTCTCAATGTTTACGCGCTTTTCAATGACATAATTTTCTCATGCTG 733
Db 430 IleglyValleuAlaSerThrpheserAsnleuValphealalyS---TrpHeme 448
QY 734 TGGATGATGATAGACCAACATACACAGTTGATATGATTAATGCTGGGCAATAAG 793
Db 449 TrpAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 464
QY 794 CAT-----GATTCTTGAGAGATGACAAAACCTTTGGCGCTTTTGGAGTGT 844
Db 465 HisthrHsAspHsPheMetLysProAlaLysValIleThrThrLeuSerValleu 484
QY 845 TTTCTTTAGGTCGTTGCTCTCGGCTTTGGAGCTCGATTCGTTTGGGAATGAT 904
Db 485 AlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 502
QY 905 TTTAAAGAGGGTATGCTTTACCTTATCCAAAAGAGCATGCGATCAGCGATGCT 964
Db 503 PheAlaProglyThrThrLeuThralaAlaAlaAlaAlaAlaAlaAlaAlaAla 520
QY 965 CAATGCGTGGCAAGTGTGATTAACATACAGAGAGCTGCTTCTT----- 1012
Db 521 GluAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 536
QY 1013 -----CTTACAGACTTCGCTATCAACATTTGGATCTTCAGAAAAGATC 1057
Db 537 GluSerAlaThrIleGluHsAspThrThrProglyGlnGlnGlnGlnGlnGln 556
QY 1058 AAAATCTTTTGAATGAAGCTTTAAGCTATACATGACAGATGAGCGCTCTCTCA 1117
Db 557 LysValProglyLeuAsnAspAla-----GluValLysGlnIleGlyAlaAla 574
QY 1118 AAATTAACATCATGAGCTGGCGTATTGTGGATTTGTTCAGAAAAGAGCGCTGAT 1177
Db 575 LysLeuPro----- 577
QY 1178 CTCTACGAAACTTAAACGAAACGAAATTTGGTCAAGGCTAAGCAGCAACTATCG 1237
Db 578 ---GlnGlyGlnValleuAlaSerGluThrValGlyProAlaValGlyLysGlu 596
QY 1238 AAGAAATACGCTTATCAGCGACCATCGGCTTTAGAGCTTTGGCAATCATCTTGC 1297

```

```

Db 597 GluThyThrIleTyAlaValleuLeuGly-----LeuGlyLeuIleLeuVal 612
QY 1298 TATGTGATTTGGCGCTTTGAATGGCAATATGCTTTCAGTCCGATAGCGCTTAATCAT 1357
Db 613 TyAlaGlyPheArgPheAspPheIleMetGlyLeuGlySerIleIleAlaAlaIleHs 632
QY 1358 GACCTTTGGCTACCTGTGACAGTCTTGTATAGACATTTCTTTTGAAGAAATTCAA 1417
Db 633 AspVal-----AlaIleAlaMetGlyLeuPheSerLeuLeuGlyLeuGlu 647
QY 1418 ATAGATTTGCAAGCATGCTGCTTTAATGATGATGATGATGATGATGATGATGAT 1477
Db 648 PheThrValAlaSerValAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 667
QY 1478 TTGATCTTTTGTATGCTATTTCTGGAAGATCGCCAGCAACCTGTTTACCCTATGAT 1537
Db 668 IleIleValSerAspArgIleArgGluAsMetLysThrMetArgGlyHsSerTyArg 687
QY 1538 GTTTTATGATATGATGCGCTTCAAAAAGACGTTGACCGGACGCTGATGACACAGTACA 1597
Db 688 GluIleValAsnAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 707
QY 1598 ACTCTATCAGTTTGTATGCTTTTGTATGAGCGGCTCTCTGCTTATATTTGCA 1657
Db 708 ThrMetLeuProLeuIleSerLeuLeuIlePheGlyGlyProValleuAlaArgAspHs 727
QY 1658 TTTATTTAGCAATGAGGATTTCTTCTGGAACCTTATGCTTCTTATTTATGCAACT 1717
Db 728 LeuLeuLeuValGlyIleLeuValGlyThrTySerSerIleTyAlaAlaPro 747
QY 1718 CTGTTGTTTATGATGCTC-----CGTAAAGAAATGCTCA 1753
Db 748 LeuValValTyTrpHegluGluTrpArgAspLysAsnArgAla 761

```

RESULT 12

ID 08YGT1

PRELIMINARY; PRT; 775 AA.

AC 08YGT1;

DT 01-MAR-2002 (TREMBlrel. 20, Created)

DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)

DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)

DE Protein translocase subunit SECD / protein translocase subunit

DE SECF.

GN BMEI1076.

OS Brucella melitensis.

OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;

OC Brucellaceae; Brucella.

OX NCBI_TaxID=29459;

RN [1]

RP SEQUENCE FROM N. A.

RC STRAIN=16M / ATCC 23456 / BIOTYPE 1;

RX MEDLINE=20020109; Pubmed=1175668;

RA DelVecchio V.G., Kapatral V., Redkar R.J., Patra G., Mujer C., Los T.,

RA Ivanova N., Anderson I., Bhattacharya A., Lykdis A., Reznik G.,

RA Jekelonski L., Larsen N., D'Souza M., Berni A., Mazur M., Golsman E.,

RA Sefkov E., Elzer P.H., Hagius S., O'Callaghan D., Letesson J.-J.,

RA Haevelorn R., Kyrides N., Overbeek R.,

RT "The genome sequence of the facultative intracellular pathogen

RT Brucella melitensis."

RL Proc. Natl. Acad. Sci. U.S.A. 99:443-448 (2002).

DR EMBL: AE009547; AAL52257.1; -.

DR InterPro: IPR000408; Reg. chr. condens.

DR InterPro: IPR003335; SecD_SecF.

DR Pfam: PF02355; SecD_SecF.1.

DR TIGRFAMs: TIGR00916; 2A0604801; 2.

DR TIGRFAMs: TIGR00966; 3A0501807; 1.

DR TIGRFAMs: TIGR01129; secD.1.

DR PROSITE: PS00626; RCCL_2; UNKNOWN_1.

KW Complete proteome.

SQ SEQUENCE 775 AA; 83880 MW; 5BE72B71588A25FE CRC64;

Alignment Scores:

Pred. No.: 7.25e-47 Length: 775

Score:	653.50	Matches:	166
Percent Similarity:	52.49%	Conservative:	140
Best Local Similarity:	28.47%	Mismatches:	204
Query Match:	19.73%	Indels:	73
DB:	16	Gaps:	15

US-09-868-987-1 (1-1864) x Q8YGT1 (1-775)

Qy	38	ACTGCTAATGGACAATATTCTGCAAAACCGTGGATGGCGTATCGCTGTGTAGTGAATGACGGT	97
Db	219	ThrAlaGlnGlyAsnGluAsnSer-----PheAlaIleValValAsnAen	233
Qy	98	TATATGTGTACAGCAGCCCTATTTTAAACGTCCTCATG--AAAATCATGCCAGTGTCTCA	154
Db	234	GlnValValSerAlaProThrValSerGlyProLeuAspThrSerGluLeuGlnIleGlu	253
Qy	155	GGGAAATTTTACCACCCAGCTGAAGTGCAGCAACTCGCTCAGATTAAAAATCTCGAGCGCAT	214
Db	254	GlyAlaPheAspLeuGlnAlaAlaAsnAenMetAlaValValLeuArgSerGlyAlaLeu	273
Qy	215	TCITTTTGTCCGAGGTTCTCAGTGAAGAGACGATCTCTTCGTATCTTCGGGAAAAACAA	274
Db	274	ProGlnAlaValThrValLeuGluGluArgThrIleAlaSerAlaLeuGlyGluAspTyr	293
Qy	275	TGTACACAGGCATATTCTCAGCATGCTGTGGCTTGGCAATGCTTATTGTTTGTATGATGAGC	333
Db	294	AlaSerAlaAlaValLeuAlaAlaLeuLeuAlaAlaLeuValValGlyLeuPheMetVal	313
Qy	335	GTATATTATAGATTGGAGGCGTCATCGCTTCGGGAGCTGTCTTCTCGAATCTTTTGGCTT	394
Db	314	LeuSerTyrGlyIleLeuGlyValIleAlaLeuValAlaLeuValValAsnIleIleIle	333
Qy	395	ATCTGGCGCAGCTCTACAGTATTGGATTTGGATGGCCACTCACCTTGTTCAGGACTCGCTGGATT	454
Db	334	LeuThrAlaValLeuSerLeuIleGlyAlaSerIleSerLeuAlaSerIleAlaGlyLeu	353
Qy	455	GTTCTTGCTATGGGATGGCGGTAGATGCAATGTTCTGTATTCGAAGAATCCGAGAG	514
Db	354	ValLeuThrIleGlyLeuAlaValAspAlaHisIleLeuIleTyrGluArgValArgGlu	373
Qy	515	GAATTTTATTGTCTCAAGCTTTAAAAAATCTGTAGAAAAAGGATATACCAAGGCTTTT	574
Db	374	AspArgArgIysGlyTyrSerValValGlnAlaMetGluSerGlyPheTyrArgAlaLeu	393
Qy	575	GGAGCCATTTTGTATCTAACTTGACTACAGTATGGCCCTCAGCACTCTTTTCTCTCTA	634
Db	394	SerThrIleValAspAlaAsnLeuThrIleLeuAlaAlaLeuValLeuPheLeuLeu	413
Qy	635	GATACAGGGCTATTAAAGGGTTTGTCTTGACATTCATTATTAGGAATTTCTCTCTCAATG	694
Db	414	GlySerGlyThrValHisGlyPheAlaLeuThrValAlaIleGlyIleGlyThrThrLeu	433
Qy	695	TTTACGGCTCTTTTCATGACTAAATTTTCTTCATGCTGTGGATGAAT-----AAGACC	748
Db	434	PheThrThrLeuThrPheThrArgLeuLeuAlaGlnTyrValArgThrAlaLysPro	453
Qy	749	CAACATACACAGTTGCATATGATGAATAAGTTCGTGGGATAAACGAT---GATTTCCTG	805
Db	454	LysGluValProLysArgArgLeuLysLeuValProThrValThrHisIleProPheMet	473
Qy	806	AGAGGATGCAAAAACTTTGGGCTGTTTCTGGAAGTGTTTTCTTTTAGTTGTC-----	859
Db	474	Arg-----LeuGlnPheValThrLeuGlyIleSerValLeuAlaCysAlaIle	489
Qy	860	-----GTTGCTCTCGGTTTGGAGCCTCGGAATTCGCTTTTGGGAATGAT	904
Db	490	ValValAlaLeuPheValAsnIleGlyPheAsn-----TyrGlyIleAsp	504
Qy	905	TTTAAAGAGGGTATGCCTTTACCTTTAATCCAAA-----GAGCATGGCATCAGCGAT	958
Db	505	PheArgGlyGlySerMetValGluLeuGlnAlaArgAsnGlyAspAlaAsnLeuGluAsp	524
Qy	959	GTTGCTCAAAATGGGTGCAAAAGTTGTGCATAAATACAGGAAGCTGGTCTTCTTCTAGA	1018

Db	525	IleAsnGlu	-----ArgLeuAlaGluLeuAsnIleAspSer---	536
Qy	1019	GACTTCGGTATTCAAACATTTGGATCTTCAGAAAAAGATCAAAATCTATTTTAGTGATAAA	1078	
Db	537	-----	-----Ala	537
Qy	1079	GCITTTAAGCTATACTAAGCAGATACGAGCCTCTCTCTAAATTAACGATCATGAGCTGG	1138	
Db	538	ArgValLeuProAlaIlySerProArgSerAlaLeuVal	554	
Qy	1139	CGTTATTGTGGG-----ATTGTTGTCAGAAACAGGCGCTAGATTTCTC	1180	
Db	555	GIncluValGlyAspAlaGluInThrValAlaValIysLeuArgGlyGluPheGlu	574	
Qy	1181	TACGGAAACTCT-----AAACGGAAACGCAAAATTTTGGTCAAAAGGTAAAGCAGCAAACTATCG	1237	
Db	575	GlnAspTyrSerPheGlnArgValAspValValGlyProThrValSerGluGlnLeuSer	594	
Qy	1238	AAGAAATGCGTTATCAGCGGACCATCGGGCTTTTAGAGCTTTGGCAATCATCTTGCTC	1297	
Db	595	ArgAlaGlyValLeuAlaValIleLeuSerLeuIleGly-----IlePheIle	610	
Qy	1298	TATGTGAGTTGCGCTTTGAATGGCAATATGCTTTTCAGTGCCTATGCGCTTTAATTCAT	1357	
Db	611	TyrValTrpPheArgPheArgTrpGlnLeuAlaLeuGlyAlaValLeuSerThrLeuHis	630	
Qy	1358	GACCTTTTGGCTACCTGTGCGTCTTGTGTTATAGCACATTTCTTTTGAAGAAAAATTCAA	1417	
Db	631	AspVal-----ValIleLeuSerGlyMetPheIleValPheArgMetGlu	645	
Qy	1418	ATAGATTTGCAAGCCATTGGTGCTTTAATGACTGATTTGGGGTATTCTAATAACAATACT	1477	
Db	646	PheAsnLeuTrpSerValAlaAlaIleLeuThrIleIleGlyTyrSerLeuAsnAspThr	665	
Qy	1478	TTGATCATTTTTCATCGTATTCGTGAAGATGCCCAAGCGAACCTGTTTACCCTTATGCAT	1537	
Db	666	ValValIleTyrAspArgValArgGluLeuAsnLeuArgArgTyrLysSerAlaProLeuPro	685	
Qy	1538	GTITTAGTTAATGATCCCTCTCAAAGAGCGTTTCAGCCCGCAGGTAAATGACACACGCTACA	1597	
Db	686	AlaIleIleAspAlaSerIleAsnGlnThrLeuSerArgThrLeuLeuThrSerPheVal	705	
Qy	1598	ACTCTATCAGTTTGTGTTAATGCTTTGTTTATAGCGCGCTCTCTGCTTTAATTTTGCA	1657	
Db	706	ThrPheLeuAlaHisValProLeuTyrAlaPheGlyGlySerGluIleArgMetPheAla	725	
Qy	1658	TTTATTATGACCATAGGAGTTCTCTTAGGAACCTTTATCGTCTCTTTATATTGACCACT	1717	
Db	726	LeuAlaLeuSerValGlyIleIleValAlaSerTyrSerSerIlePheIleAlaAlaPro	745	
Qy	1718	CTGTTGTTG 1726		
Db	746	LeuLeuVal 748		

RESULT 13

REC'D 13
Q9ZBS8
ID Q9ZBS8 PRELIMINARY; PRT; 795 AA.

Q24550;
AC 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DT DT
DE Putative SecDF protein-export membrane protein.
DE
GN SCO6160 OR SC1A9.24C.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces
OC NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Saunders D.C., Harris D.;
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.


```
QY 1403 TTGAAGAAATTCAAATAGATTGCAAGCCATTGGTCTCTTAATGACTGTATTCGGGTAT 1462
Db 645 TrpLeuGlyLysProIleAspGlyValPheLeuAlaLeuLeuThrValIleGlyTyr 664
QY 1463 TCATTAAACAATACCTTGATCATTTTGGATCGTATTCTGTGAA-----GATCGC 1510
Db 665 SerValAsnAspSerValValLeuPheAspArgIleArgGluLeuLeuGlyLysGluArg 684
QY 1511 CAAGCGAACCTGTTTACCCTATGTCATGTTTATAGTTTAAATGATGCCCTTCAAAGACGTTTC 1570
Db 685 LysAla-----ProPheAspArgLeuThrAsnAspAlaIleLeuGlnThrLeu 700
QY 1571 AGCGCACGCTATGACCAACAGCTACACTCTATCAGTTTGTAAAGCTTTGTTGTTTATA 1630
Db 701 ProArgThrValAsnThrGlyMetGlyAlaValLeuIleLeuAlaSerLeuAlaIleLeu 720
QY 1631 GGCGGCTCTGCTGCTGCTTAAATTTGTCATTTATATGACCATAGGATTTCTTAGGAAC 1690
Db 721 AlaAspSerLeuThrAspPheAlaLeuAlaLeuIleGlyValGlyValGlyThr 740
QY 1691 TTATCGCTCTTTATATGACCACTCTGTTGTTGTTTATGTCCTGCGTAAGAAATCGC 1750
Db 741 TyrSerSerValPheThrAlaSerProLeuAlaIleGluLeuHisAsnArgAspThrGly 760
QY 1751 TCAAAA 1756
Db 761 SerArg 762

RESULT 14
Q92B17 PRELIMINARY; PRT; 754 AA.
AC Q92B17;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DE 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical protein lin1562.
GN LIN1562.
OS Listeria innocua.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Listeriaceae; Listeria.
OX NCBI_TaxID=1642;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CLIP 11262 / SEROVAR 6A;
RX PubMed=11679669;
RA Glaser P., Frangeul L., Buchrieser C., Ruenik C., Amend A.,
RA Baquero P., Berche P., Bloeker H., Brandt P., Chakraborty T.,
RA Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,
RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA Jones L.-M., Kaerst U., Kref J., Kuhn M., Kunst F., Kurapkat G.,
RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
RA Nordiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
RA Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,
RA Vazquez-Boland J.-A., Voss H., Wehlund J., Cossart P.;
RT "Comparative genomics of Listeria species.";
RL Science 294:849-852(2001).
DR EMBL; AL596169; CAC96793.1; -
DR L181181; LIN01562; -
DR InterPro; IPR003335; SecD_SecF.
DR Pfam; PF02355; SecD_SecF; 1.
DR TIGRFAMs; TIGR00916; 2A0604s01; 2.
DR TIGRFAMs; TIGR01129; secD; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 754 AA; 82546 MW; 047630364532DD4E CRC64;
```

Alignment Scores:

Pred. No.:	4,52e-44	Length:	754
Score:	620.50	Matches:	162
Percent Similarity:	47.13%	Conservative:	117
Best Local Similarity:	27.36%	Mismatches:	222
Query Match:	18.73%	Indels:	91

```
DB: 16 Gaps: 13
US-09-868-987-1 (1-1864) x Q92B17 (1-754)
QY 98 TATATGGTCAGCAGCCCTATTTTAAACGCTCCCATTTGAAATCATGCCAGTGCTCAGGG 157
Db 196 TyrLeuSerAlaProAsnValSerSerValLeuAspThrAspLysValGluIleSerGly 215
QY 158 AATTTATCCACCGCTGAAGTGAACAACTCGCTCAGATTTTAAATCTTGAGCGATGCT 217
Db 216 SerPheThrThrGluGluAlaLysAspLeuAlaGluLeuLeuAsnSerGlyAlaLeuPro 235
QY 218 TTTGTTCCCGAGGTTCTCAGTGAAGAGCAGTCTCTCTGATCTTGGGAAAAACAATGT 277
Db 236 ValLysMetThrGluValTyrSerThrSerValGlyAlaGlnPheGlyGlnAspAlaLeu 255
QY 278 ACACAGGCAATTTCTCAGCATCTGTGGCTTGGCAATGCTTATTGTTTGGATGAGCGTA 337
Db 256 GlnGluThrIleLeuAlaGlyIleIleGlyValMetAlaIlePheIlePheMetMetAla 275
QY 338 TATTATAGATTTTGGAGCGTTCATCGCTTGGGAGCTGTTCTTCTGAATCTTTTCTTATC 397
Db 276 ValTyrArgLeuProGlyValIleAlaSerIleThrLeuValAlaTyrThrTyrLeuVal 295
QY 398 TGGGCACTCTACAGTATTGAGTGGCCCACTCACCCTGTGACGAGCTCGCTGGGATGTT 457
Db 296 LeuLeuIleLeuSerLeuLeuAsnAlaThrLeuThrLeuProGlyIleAlaGlyLeuIle 315
QY 458 CTGTGCTATGGGATGGCGTAGATGCAATGTTCTTGTATTCGAAAGAATCCGAGAGGAA 517
Db 316 LeuGlyIleGlyMetAlaValAlaAspAlaAsnValIleThrTyrGluArgIleLysGluGlu 335
QY 518 TTTTATTTGCTCAAAAGTCTTAAAAATCTGTAGAAAAAGGATATACCAAGGCTTTTGGGA 577
Db 336 IleLysValGlyArgSerThrLysAlaAlaPheGluValGlyLysGluAlaPheArg 355
QY 578 GCCATTTTGAATTTCACTAGCTACAGTATGCGCTCAGCAGCTCTTCTTCTTCTCTAGAT 637
Db 356 AlaIleLeuAspGlyAsnLeuThrThrLeuIleValAlaAlaValLeuPheTyrPheGly 375
QY 638 ACAGGCGCTATTAAAGGTTGCTTTGACATTCATTTTACGAATTTTCTCTCAATGTTT 697
Db 376 ThrSerSerIleLysGlyPheAlaThrValLeuIleIleSerIleLeuValSerPheLeu 395
QY 698 ACGGCTCTTTTCATGACTAAATTTTCTTCTCATGCTG-----TGAGTAAT 742
Db 396 ThrAlaValTrpGlySerArgPheLeuLeuGlyLeuValLysSerAsnTrpLeuAsn 415
QY 743 AAG-----ACCAACATACACAGTTGCGATGATGATGAATAAGTTCGTG 784
Db 416 AsnLysProGlyPhePheAlaValLysArgLysAspIleHisAsnLeuHisGluGlyIle 435
QY 785 GGATA-----AAGCATGATTTCTTGAGAGGATGCAAAAAA 820
Db 436 AsnSerPheSerLeuLysThrHisPheAspArgTyrAspPheValLysHisArgLeu 455
QY 821 CTTTGGGCTGTTTCTGGAAGTGTCTTTTAGTGTGCGTTCCTCGGGGTTTGGAGCC 880
Db 456 PheLeuSerIlePheAlaAlaIleValIleValIleValIleLeuSerIlePheLys 475
QY 881 TGAATTTCCGCTTTTGGGAATGGAATTTTAAAGAGGGGATGCGCTTTTACCTTTTAAATCCAAA 940
Db 476 LeuAsn-----LeuGlyIleAspPheAlaSerGly-----ThrArg 487
QY 941 GAGCATGGCATCAGCATGTTGCTCAATCGCTGGCAAAAGTTGTGCATAAATACATACAGAA 1000
Db 488 AlaGluValThrAlaAsnGlnThrLeuThrGluThrGlnIleLysLysAspLeuAspAla 507
QY 1001 GCTGGTCTTTCTTAGAGACTTCCCGTATTCAAACATTTTGGATCTTCA----- 1048
Db 508 IleAspMetProSerAsnAspIleValPheGlnGlySerGlySerLysThrAlaValVal 527
QY 1049 -----GAAAAAGATCAAAATCTATTTTAGTCAT 1075
```

```

Db 528 SerTyrIysGlyThrLeuSerGlnAsnAspValAlaIlysPheIysAspTyrPheGluAsp 547
Qy 1076 AAAGCTTAACTATACAGAGATACAGAGCTCTCTCTAAATTAACATCATGAGC 1135
Db 548 Lys-----TyrLysHisGluProSerIleSerThrValSerProThrVal----- 562
Qy 1136 TGGCGTTATTGGGATGTGTGTCTGAAAGAGAGCTTAAGTTCTCTAGGAAACTTAA 1195
Db 563 -----GlyLysGluLeu 566
Qy 1196 CGAAACGAAATTTGGTCAAGGTAAAGACAAACTATCGAAGAAATGCGTATACG 1255
Db 567 AlAllysAsnGlyPheTrpAla----- 573
Qy 1256 GCGACCATCGGGCTTTTAGA---GCTTTGGCAATCATCTGCTCATGTAGATTGGCGC 1312
Db 574 -----LeuGlyValAlaSerValLeuIleValLeuTyrIleAlaValArg 588
Qy 1313 TTGGAATGCGCAATATGCTTTACAGTCCGATGCGCTTAAATTCATGACCTTTGGCTACC 1372
Db 589 PheGluPheTyrMetGlyIleAlaIleAlaLeuSerLeuLeuPheAsp----- 604
Qy 1373 TGTGAGCTTGTATTAGCACATTTCTTTGAAGAAATTCATGATGATTGGCAAGCC 1432
Db 605 ---AlaPheIleIlePheIlePhePheSerValThrArgLeuGluValAlaSpheThrPhe 623
Qy 1433 ATTGGTCTTAAATGATCTGTATTGGGGTATTTCATTAACAATCTTATGATCTTTGAT 1492
Db 624 IleAlaIleAlaValLeuThrValIleGlyTyrSerIleAsnAspThrIleValThrAlaAsp 643
Qy 1493 CGTATTCGTGAA---GATCGCCAGACGAACTGTTACCCCTATGCAATGTTTA----- 1543
Db 644 ArgIleAspAspIleSerMetLysMetGlnArgPheLysTrpLysGluGluIleAlaAsp 663
Qy 1544 ---GTTATGATGACCTTCNAAGAGCTTCAGCCGATGATGATGATGATGATGATGAT 1600
Db 664 AlAlaValAsnLysAlaLeuAspGlnThrPheThrArgSerIleAsnThrIleLeuThrVal 683
Qy 1601 CTATCAGTTTGTATTGATGCTTTTGTATTAGCGGCTCTCTGCTTATTTTGGCATTT 1660
Db 684 IlePheThrValLeuAlaValValLeuPheGlySerGluSerIleLeuAsnPheSerIle 703
Qy 1661 ATTATGACCATTAAGGATCTTCTTACGAACTTATGCTCTTATTAATTCACACCTCTG 1720
Db 704 AlaLeuLeuValGlyLeuValSerSerValPheSerSerIlePheMetAlaMetGlnLeu 723
Qy 1721 TTGTTGTTTATGTCGCTTAAGAAATGCTCAAAA 1756
Db 724 TrpTyrValPheLysAlaArgGlnLeuArgLysLys 735

RESULT 15
ID 08Y703 PRELIMINARY; PRF: 754 AA.
AC 08Y703;
DT 01-MAR-2002 (TReMBLrel. 20, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Hypothetical protein Imol527.
GN Imol527.
OS Listeria monocytogenes.
OC Bacteria; Firmicutes; Bacillales; Clostridium group; Bacillales;
OC Listeriaceae; Listeria.
OX NCBI_TaxID=1639;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EGD-E / SEROVAR 1/2a;
RX MEDLINE=21537279; PubMed=11679669;
RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
RA Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,
RA Charbit A., Cherouani F., Couve E., de Daruvar A., Deboux P.,
RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA Entian K.-D., Fshni H., Garcia-del Portillo F., Garrido P.,

```

```

RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA Jones L.-M., Kaerster U., Kreft J., Kuhn M., Kunst F., Kurapack G.,
RA Madueno E., Maitounam A., Mata Vicente J., Ng E., Nedjati H.,
RA Nordstok G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
RA Rammel B., Rose M., Schlueter T., Simons N., Tierrez A.,
RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.,
RT "Comparative genomics of Listeria species.";
RL Science 294:845-852 (2001).
DR EMBL; AL591979; CAC9605.1; -.
DR ListIst; LM001527; -.
DR InterPro; IPR003335; SecD_SecF.
DR Pfam; PF02355; SecD_SecF.1.
DR TIGRFAMs; TIGR00916; 2A0604s01; 2.
DR TIGRFAMs; TIGR01129; secD; 1.
KM Hypothetical protein; Complete proteome.
SQ SEQUENCE 754 AA; 82614 MW; F2BA63B501EB58B CRC64;

Alignment Scores:
Pred. No.: 6,68e-44 Length: 754
Score: 618.50 Matches: 166
Percent Similarity: 46.79% Conservative: 111
Best Local Similarity: 28.04% Mismatches: 224
Query Match: 18.67% Indels: 91
DB: Gaps: 13

US-09-868-987-1 (1-1864) x 08Y703 (1-754)
Qy 98 TATATGTCAGACGCCCTATTTAAACGTCCTCAATGAAATCATGCCAGTCTCAGG 157
Db 196 TyrLeuSerAlaProAsnValSerSerValLeuAspThrAspLysValGluIleSerGly 215
Qy 158 AAATTTCACCCGTCGAAGAGAGCAAACTGCCCTCAATTTAAATTTGAGCATGCT 217
Db 216 SerPheThrArgGluAlaLysAspLeuAlaIleLeuLeuAsnSerGlyAlaLeuPro 235
Qy 218 TTGTCCCGAGGTTCTCAGTGAAGAGCATCTCTTGATCTTGGAAAAACAATGT 277
Db 236 ValLysMetLysGluValIlyrSerThrSerValGlyAlaGlnPheGlyGlnAspAlaLeu 255
Qy 278 ACACAGGACATTAATCTCAGATGCTGTGCTTGGCAATGCTTATGTTGATGAGCGTA 337
Db 256 GlnGluThrIleLeuAlaGlyIleIleGlyValIleAlaIlePheIlePheMetMetAla 275
Qy 338 TATATGATTTGAGAGCGCTCATGCTTGGGAGCTTCTCTGATCTTTGCTATC 397
Db 276 ValTyrArgLeuProGlyValIleAlaSerIleThrLeuValAlaTyrThrTyrLeuVal 295
Qy 398 TGGCAGCTCTACAGATTTGATGATGCGCACCTCACTTTCAGAGACTCGCTGGATTGT 457
Db 296 LeuLeuIleLeuSerLeuLeuAsnAlaThrLeuThrLeuProGlyIleAlaGlyLeuIle 315
Qy 458 CTTCATGAGGATGCGCGTAGATGCAATGTTCTTATTCGAAGATCCGAGAGAA 517
Db 316 LeuGlyIleGlyMetAlaValAlaAspAlaValIleThrTyrGluArgIleLysGluLys 335
Qy 518 TTTTATGTCTCAAACTTTAAATCTGTAGAAAAGATATACCAAGCTTTTGA 577
Db 336 IleLysValGlyArgSerThrLysAlaAlaPheGluValGlyLysGluAlaPheArg 355
Qy 578 GCCATTTTGTATTAAGTCTACAGATGATGCGCTTGGCTCAGAGACTTTTCTCTGAT 637
Db 356 AlaIleLeuAspGlyAsnLeuThrThrLeuIleValAlaAlaValAlaPheTyrPheGly 375
Qy 638 ACAGGCTTATTAAGGATGCTTTCATGATGATGATGATGATGATGATGATGATGAT 697
Db 376 ThrSerSerIleLysGlyPheAlaThrValLeuIleIleSerIleLeuValSerPheLeu 395
Qy 698 AGCGCTCTTTCATGATCAATTTTCTTCATGCTG-----TGATGAT 742
Db 396 ThrIleValTrpGlySerArgPheLeuGlyLeuLeuValLysSerAsnTrpLeuAsn 415
Qy 743 -----AAGCAACCAATACAGAGTTGCAATGATG 772

```


QY 623 CTTTCTCTAGATACAGGCGCTTAAGGTTGCTTGAATTTAGATTTAGAAAT 682
 |||||
 Db 456 LeuPheTYrPheGlyThrIleuIleArgIlePheSerIleSerIleuIleuSerIle 475
 QY 683 TTCCTTCAATGTTTACGGCTCTTTTCATGACTAAATTT----- 721
 |||||
 Db 476 IlePheIleuPheValMetIleuIleIleThrArgMetAlaThrSerIleuIleGly 495
 QY 722 -----TTCCTTCATGCTGTGATGAATAAGCCCAACATACACAGTTCATATGAT 775
 |||||
 Db 496 ThrGlyPhePhe-----AspAsnArgLeuTYrIleu----- 505
 QY 776 AAGTTCGTGGGATTAACACATGATTTCTTGAGAGATGCAAAAACTT----- 823
 |||||
 Db 506 -----IleGlyValGlyArgSerPhePheThrIleGlyIleIleIleThrPhePhe 523
 QY 824 -----TGGCGCTGT-----TCTGGAAGTGTCTTCT 850
 |||||
 Db 524 GluIleuAsnTYrPheIleuSerIleTYrPheIleIleuPheGlyValPheVal 543
 QY 851 TTAGGTGCGTGTCTCTCGGCTTGGAGCTGGAATTCGCTTTGGGA----- 898
 |||||
 Db 544 LeuIleAlaIleIleIleIleIleIleIleIleIleIleIleIleIleIleIleIle 562
 QY 899 -----ATGATTTTAAAGAGGCTATGCTTACCTTTAATCCAAA--- 940
 |||||
 Db 563 GlyPheGlyArgSerLeuGluPheGlnGlyThrAsnIleSerPheGluIleuSerSer 582
 QY 941 -----GAGCATGCGATGCGATGCTGCTCAATGCGGCAAGTGTGCATTAACCTA 994
 |||||
 Db 583 IleAspAsnLeuSerLeuAlaGluAlaGlnGluIleIleIleIleIleValAsnLeuIle 602
 QY 995 CAGGAAGCTGCTCTTCTTCTTGAAGACTCCGATTCACAACATTTGGATCTTCAGAAAAG 1054
 |||||
 Db 603 AsnThrAlaAsn-----GlySerGluAsnLeu 611
 QY 1055 ATCAAAATCTATTTTGAATGAATTAAGCTTAAAGCTTAAAGATACAGACTCTCTC 1114
 |||||
 Db 612 IleThrSerGlyAlaGlnAspIlePheIleAsnAlaIleAsnGlnAlaIleAsnSer 630
 QY 1115 CTAAATTAAGATCATGATGACGCGGCTTATGTGGATGTTGTCAGAAACGCGCTGGA 1174
 |||||
 Db 631 -----TYrValValSerLeuGlnThrIle 638
 QY 1175 TTTCTCTACGAAACTTAACGAAC-----GCAAAATTTTGCTCA 1216
 |||||
 Db 639 GlnValPheAspGlnAspGlnIleAsnTYrIleIleGlnGlnValSerIleThrPheSer 658
 QY 1217 AAGTAACT-----AGCAAACTATCGAAGAA 1243
 |||||
 Db 659 AsnIleSerSerThrPhePheIleValSerThrAsnGlnAlaSerSerIleValIleAsn 678
 QY 1244 ATGCGCTTATCAGGACACATGCGGCTTATGAGCTTGGCAATCATCTGTGCTATG 1303
 |||||
 Db 679 SerIleValSerIleSerIleGlyIleValGly-----IleValLeuTYrThr 694
 QY 1304 AGTTTGCCCTTGAATGCAATATGCTTCAAGTCCGATGCGCTTAAATCATGACCTT 1363
 |||||
 Db 695 IlePheArgLeuIleTYrThrPheSerIleAlaIleValAlaIleuIleuIleAspIle 714
 QY 1364 TTGGCTACCTGTGACGCTTGTATTAAGACATTTCTTTGAAGAAAATTAATAGAT 1423
 |||||
 Db 715 -----ValPheValIleAlaPhePheIleIlePheArgLeuIleuIleSer 729
 QY 1424 TTGCAAGCATTTGCTTATTAAGTACTGATGCGGATTTCTTAACAATCTTGTATC 1483
 |||||
 Db 730 ProIlePheValAlaIleIleuSerIleIleGlyTYrSerIleAsnAspThrIleVal 749
 QY 1484 ATTTTGATCGATTCGATGACATGCGCAAGCAACCTGTTT-----ACCCCTATG 1534
 |||||
 Db 750 ThrPheAspArgIleIleuSerIleIleMetIleSerGluTYrTYrIleGlnIleuIleAspGln 769

QY 1535 CATGTTTAACT-----AATGATGCCCTTCAAAAAGAGCTTACCGACGCGATATG 1585
 |||||
 Db 770 LysValIleuIleAsnIleThrAsnIleAlaIleIleAspThrIleLysArgSerMetLeu 789
 QY 1586 ACAACAGCTCAACACTATCATGTTTGTGTAATGCTTTTGTATAGCGGCTCTCTGTC 1645
 |||||
 Db 790 ThrThrIleThrIleThrIleValThrIleValIleuIleuSerPheArgIleSerThrGlu 809
 QY 1646 TTTAATTTTGCATTTATTAAGCATTGAGCATGAGGATCTTTAGAACTTTATGCTCTTAT 1705
 |||||
 Db 810 ValAsnIleAsnIleAlaIleuPheValGlyLeuIleSerGlyThrTYrSerSerIlePhe 829
 QY 1706 ATGCA 1711
 |||||
 Db 830 IleAla 831
 RESULT 17
 Q8RFJ6
 ID Q8RFJ6 PRELIMINARY; PRT; 411 AA.
 AC Q8RFJ6;
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
 DE Protein translocase subunit secD.
 GN FNO699.
 OS Fusobacterium nucleatum (subsp. nucleatum).
 OC Bacteria; Fusobacteriia; Fusobacterium.
 OX NCBI_TaxID=76856;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 25586;
 RX MEDLINE=21886394; PubMed=11889109;
 RA Kapralat V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A.,
 RA Bhattacharya A., Bateman A., Gardner W., Greshkin G., Zhu L.,
 RA Vasileva O., Chu L., Kogan Y., Chaga O., Goldsman E., Bernal A.,
 RA Larsen N., D'Souza M., Malinas T., Pusch G., Haselkorn R.,
 RA Fongstein M., Kyridis N., Overbeek R.;
 RT "Genome sequence and analysis of the oral bacterium Fusobacterium
 RT nucleatum strain ATCC 25586.";
 RL J. Bacteriol. 184:2005-2018 (2002).
 DR EMBL; AEO10580; AAL94895.1; -;
 KW Complete proteome.
 SQ SEQUENCE 411 AA; 44125 MW; D754D2B02035235F CRC64;
 Alignment Scores:
 Pred. No.: 1,54e-30 Length: 411
 Score: 460.50 Matches: 89
 Percent Similarity: 63.60% Conservative: 56
 Best Local Similarity: 39.04% Mismatches: 82
 Query Match: 13.90% Indels: 1
 Gaps: 1
 US-09-868-987-1 (1-1864) x Q8RFJ6 (1-411)
 QY 62 AACGCTGATGCGGCTATGCTAGTATGAGCGTTATATGTCAGACGCTATTTTA 121
 |||||
 Db 166 AsnIleGlyArgGlnLeuAlaIleThrLeuAspGlyValGlnIleAlaProIleIle 185
 QY 122 AACGCTCCATGAAAATCAATGCAAGTGC-----TCAAGGAATTTTACCAACCGTGAAGT 178
 |||||
 Db 186 AsnThrGluIleSerGlyGlySerGlyValIleThrGlyAsnTYrThrValGlnGlnAla 205
 QY 179 AGCAAACTGCGCTCAGATTTTAAATCTGAGCGGATCTTTGTTCCGAGGTTCTCAGT 238
 |||||
 Db 206 LysGlyThrAlaIleAlaLeuLeuAsnIleGlyAlaLeuProIleLysAlaGlnIleAlaGlu 225
 QY 239 GAAGAAGCATCTTCTGATCTTGGAAAACAAATGATACAAAGCATATCTACGA 298
 |||||
 Db 226 ThrArgThrValGlyAlaThrLeuGlyAspIleSerIleAlaGlnSerLysAsnIleGly 245
 QY 299 TGCTGTGCTGCGCAATGCTTATGTTTGTGTAAGCGCTATATTAATTTGAGAGCGCTC 358
 |||||
 Db 246 MetValAlaIleValLeuIleIleThrValPheMetIleIlePheTYrArgLeuProGlyIle 265

```

Qy 359 ATCCGCTCGGAGCTGTTCTCTGAATCTTTGCTATCTCGGAGCTCTACAGTATTG 418
    |||||
Db 266 ILeAlaAspLeuAlaIlePheGlyPheIleThrPheAlaCysLeuAsnPheIle 285
    |||||
Qy 419 GATCGGCACCTCAGCTGCTAGGACTCGCTGGATTGTTCTGCTATGGGATGCGCGTA 478
    |||||
Db 286 AspAlaThrLeuThrLeuProGlyIleAlaGlyPheIleLeuSerLeuGlyMetAlaVal 305
    |||||
Qy 479 GATCAAAATGTTCTGATTTCCAAAGATCCGAGAGCAATTTTATTGTCrCAAAAGTCTT 538
    |||||
Db 306 AspAlaAsnValIleIlePheGluArgIleLysGluLeuLeuArgPheGlyAsnSerIle 325
    |||||
Qy 539 AAAAAATCTAGAAAAAGATATACCAAGCTTTTGGAGCCATTTTGTATCTTAACCTTG 598
    |||||
Db 326 ArgAsnSerIleAspSerGlyPheAsnLysGlyPheIleAlaIlePheAspSerAsnLeu 345
    |||||
Qy 599 ACTACAGTATTGGCTCAGCACTCTCTTCTCTCTAGATACAGGCGCTATTAAAGGCTTT 658
    |||||
Db 346 ThrThrLeuIleIleThrIleLeuPheValPheGlyThrGlyProIleIleGlyPhe 365
    |||||
Qy 659 GCTTTGACATGATTTTAGGAATTTCTCTCAATGTTTACCGCTCTTTTCATGACTAAA 718
    |||||
Db 366 AlaValThrLeuAlaLeuGlyThrLeuAlaSerMetPheThrAlaIleThrValThrLys 385
    |||||
Qy 719 TTTTCTTCATGCTGTGATGAAT 742
    |||||
Db 386 ValLeuLeuLeuThrPheValAsn 393
    |||||

RESULT 18
Q820H5 ID Q820H5 PRELIMINARY; PRT; 474 AA.
AC Q820H5;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Protein-export membrane protein.
GN SEC OR ALL0121.
OS Anabaena sp. (strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=103690;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21595285; PubMed=11759840;
RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
RA Watanabe A., Iriguchi M., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakazaki N., Shimpo S., Sugimoto M., Takazawa M., Yamada M.,
RA Yasuda M., Tabata S.;
RT "Complete genomic sequence of the filamentous nitrogen-fixing
RT cyanobacterium Anabaena sp. strain PCC 7120.";
RL DNA Res. 8:205-213(2001).
DR EMBL: AP003581; BAB7645.1; -
DR InterPro; IPR000911; Ribosomal_L11.
DR InterPro; IPR003335; Secd_Secf.
DR Pfam; PF02355; Secd_Secf.1.
DR TIGRFAMs; TIGR00916; 2A0604s01.1.
DR TIGRFAMs; TIGR01129; secD.1.
DR PROSITE; PS00359; RIBOSOMAL_L11; UNKNOWN_1.
KW Complete proteome.
SQ SEQUENCE 474 AA; 50454 MW; 67ABEB4BE4A3D564 CRC64;

Alignment Scores:
Pred. No.: 1.96e-28 Length: 474
Score: 436.00 Matches: 89
Percent Similarity: 60.61% Conservative: 51
Best Local Similarity: 38.53% Mismatches: 85
Query Match: 13.16% Indels: 6
DB: 16 Gaps: 1

US-09-868-987-1 (1-1864) x Q820H5 (1-474)

```

```

Qy 59 GCAAACCGTGGATGGCTATGGCTGTAGTATTGACGGTTATATGGTCAGCAGCCCTATT 118

```

```

Db 214 AlaGlyThrGlyArgSerIleGlyIlePheLeuAspAsnGluLeuIleSerAlaProAsn 233
    |||||
Qy 119 TTAACCTCCCATGAAAAAT-----CATGCCAGTGTCTCAGGAAA 160
    |||||
Db 234 ValGlyIleGluPheAlaSerThrGlyIleThrGlySerAlaValIleThrGlyArg 253
    |||||
Qy 161 TTATCCACCCTGAGTGAACAACTCGCTCAGATTAAATCTCGAGCGATCTCTTTT 220
    |||||
Db 254 PheThrAlaGlnGlnAlaAsnAspLeuGlyValGlnLeuArgGlyIleAlaLeuProVal 273
    |||||
Qy 221 GTTCCCAGAGTTCTCAGTGAAGAGACCATCTCTCTCATCTTGGGAAAAACAATGTACA 280
    |||||
Db 274 ProValGluIleAlaGluIleArgThrValGlyAlaThrLeuGlyLysAspSerIleAsn 293
    |||||
Qy 281 CAAGGCATTATCTCAGCATGCTGTGGCTTGCATGCTATTGTTTGTATGAGCGGTATAT 340
    |||||
Db 294 SerSerIleThrAlaGlyLeuGlyGlyLeuThrLeuValLeuIlePheMetValValTyr 313
    |||||
Qy 341 TATAGATTTGGAGCGTCATCGCTTCGGAGCTGTTCTCTGAATCTTTTCTCTATCTGG 400
    |||||
Db 314 TyrArgLeuProGlyLeuIleAlaAspIleSerLeuIleIleTyrAlaIleLeuThrTrp 333
    |||||
Qy 401 GCAGCTCTACAGTATTGGATGGCCACTCACCTTGTCAAGCTCGCTGGGATGTTCTT 460
    |||||
Db 334 AlaSerPheAlaLeuLeuGlyIleThrLeuThrLeuProGlyIleAlaGlyPheIleLeu 353
    |||||
Qy 461 GCTATGGGATGGCCGTAGATGCAAAATGTTCTGTATTTCGAAAGAAATCCGAGAGGAATTT 520
    |||||
Db 354 SerIleGlyMetAlaValAlaAsnValLeuIlePheGluArgThrArgGluGluLeu 373
    |||||
Qy 521 TTATGTCTCAAGTCTTAAAAAATCTGTAGAAAAAGATATACCAAGGCTTTTGGAGCC 580
    |||||
Db 374 GlnAlaGlyLysSerLeuTyrArgSerValGluSerGlyPheTyrArgAlaPheSerSer 393
    |||||
Qy 581 ATTTTGTATCTAATCTAGTACTACAGTATTGGCTCAGCACTCTCTTTCTCTCAGATACA 640
    |||||
Db 394 IleLeuAspGlyAsnValThrValIleAlaCysAlaAlaLeuPheTrpLeuGlyAla 413
    |||||
Qy 641 GGGCTATTAAAGGTTTGTCTGACATGATTTTGTAGAAATTTTCTTCAATGTTTACG 700
    |||||
Db 414 GlyLeuValLysGlyPheAlaLeuThrLeuAlaLeuGlyValAlaValSerMetPheSer 433
    |||||
Qy 701 GCTCTTTTCATGACTAAATTTTCTTCTCATGCTG 733
    |||||
Db 434 AlaValThrCysSerArgThrLeuMetPheLeu 444
    |||||

RESULT 19
Q9HX11 ID Q9HX11 PRELIMINARY; PRT; 620 AA.
AC Q9HX11;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Secretion protein SecD.
GN SEC OR PA3821.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrenner P.,
RA Hickey M.J., Brinkman F.S.L., Huftagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltzy L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen.";
RL Nature 406:959-964(2000).

```

DR EMBL: AE004799; AAC07208.1; -.
 DR InterPro: IPR003335; SecD_SecF.
 DR Pfam: PF02355; SecD_SecF; 1.
 DR TIGRFAMs: TIGR00916; 2A0604s01; 1.
 DR TIGRFAMs: TIGR01129; secD; 1.
 KM Complete proteome.
 SQ SEQUENCE 620 AA; 67674 MW; BA8740DB46A099FD CRC64;

Alignment Scores:
 Pred. No.: 1.98e-27 Length: 620
 Score: 424.00 Matches: 103
 Percent Similarity: 53.71% Conservative: 49
 Best Local Similarity: 36.40% Mismatches: 83
 Query Match: 12.80% Indels: 48
 DB: 16 Gaps: 5

US-09-868-987-1 (1-1864) x Q9FX11 (1-620)

QY 62 AACGGTGGATGGCGTATGCTGATG----- 88
 DB 357 AsnValGlyArgSerMetAlaValAlaPheIleGlnLysProValThrArgTyrThr 376
 QY 89 -----ATTGACGGT----- 97
 DB 377 LysGlnMetValAspGlyValGlnLysGluValAlaValProAlaPheLysGluLys 396
 QY 98 TATATGCTACGACGCCCTATTTTAAAGTCCCATTTGAAAAATCATGCCAGTCTCAGGG 157
 DB 397 GlnIleIleSerLeuAlaThrIleGlnSerProLeuGlyAsnGlnPheArgIleThrGly 416
 QY 158 AAATTTACCAACCGTGAAGGAGCAACATCGCTCAGATTAAATTTGAGCGCATGCT 217
 DB 417 LeuAspGlyProGlyGlnSerSerGlnLeuAlaLeuLeuAlaGlyGlyLeuAla 436
 QY 218 TTTGTTCCCGAGTTCCTCAAGAGACAGATCTCTCTGATCTTGGGAAAAACAATGT 277
 DB 437 AlaProMetCyrPheAlaGluArgThrIleGlyProSerLeuGlyAlaAspAsnIle 456
 QY 278 ACACAGGCAATTTATCTCAGATGCTGTCGCGCATTCATTTGTTGATGAGCGTA 337
 DB 457 AlaLysGlyIleAspAlaSerLeuThrPrgLysMetLeuPheValSerLeuPheIleIleVal 476
 QY 338 TATATGATTTGGAGGGGTCATGCTGCGGAGCTGTTCTTCGATCTTGGGATTC 397
 DB 477 IleTyrArgPhePheGlyValIleAlaThrValAlaLeuAlaPheAsnMetValMetLeu 496
 QY 398 TGGGACGCTACAGATATTTGGATGCGGCACCTGTCAGAGCTGCTGGATTTG 457
 DB 497 ValAlaLeuMetSerIleLeuGlyAlaThrLeuThrLeuProGlyIleAlaGlyIleVal 516
 QY 458 CTTCGCTATGGGATGCGCGTACGATGCAAAATGTTCTTGTATTCGAAAATCCGAGAGGA 517
 DB 517 LeuThrMetGlyMetAlaValAspAlaAsnValIleuIlePheSerArgIleArgGlnLys 536
 QY 518 TTTTATGTCTCAAGCTTAAAGCTGTAAGAAATCTGTGAAAAAGATTAACAAGGCTTTGA 577
 DB 537 LeuAlaAsnGlyMetSerValGlnArgAlaIleHisGlnGlyPheAsnArgAlaPheThr 556
 QY 578 GCCATTTTGGATTTCACTTGACTACAGATTTGGCGCTCAGACCTTTTCTTCTAGAT 637
 DB 557 AlaIleLeuAspAlaAsnLeuThrSerLeuLeuValGlyGlyIleLeuThrAlaMetGly 576
 QY 638 ACAGGGCTATTTAAAGGTTGCTTGAACATTTAGATTTTGAAGATTTTCTCTTCAATGTTT 697
 DB 577 ThrGlyProValLysGlyPheAlaValThrMetSerLeuGlyIleIleThrSerMetPhe 596
 QY 698 ACGGCTCTTTTCATGACCTAAATTTTCTTCACTGCTGTGAGTAAGTAACCAACATATCA 757
 DB 597 ThrAlaIleMetValThrArg----- 603
 QY 758 CAGTTGATGATGATGAATAGTTCTGGGATTAAGCATGATTTCTTGAGAGATGACAA 817
 DB 604 -----AlaMetValAsnLeuIlePheGly---GlyArgAspPhe-----Lys 616

QY 818 AAATTTGG 826
 DB 617 LysLeuThr 619

RESULT 20

ID Q8XJ19 PRELIMINARY; PRT; 419 AA.
 AC Q8XJ19;
 DT 01-MAR-2002 (TREMBlrel, 20, Created)
 DT 01-MAR-2002 (TREMBlrel, 20, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel, 21, Last annotation update)
 DE Protein-export membrane protein.
 GN SECY OR CPE1942.
 OS Clostridium perfringens.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
 OC Clostridiales; Clostridiaceae; Clostridium.
 NCBI_TaxID=1502;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=13 / TYPE A;
 RX PubMed=11792842;
 RA Shimizu T., Ohtani K., Hirakawa H., Ohshima K., Yamashita A.,
 RA Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.,
 RT "Complete genome sequence of Clostridium perfringens, an anaerobic
 RT flesh-eater";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:996-1001 (2002).
 DR EMBL: AP003192; BAB81648.1; -.
 DR InterPro: IPR003335; SecD_SecF.
 DR Pfam: PF02355; SecD_SecF; 1.
 DR TIGRFAMs: TIGR00916; 2A0604s01; 1.
 DR TIGRFAMs: TIGR01129; secD; 1.
 KM Complete proteome.
 SQ SEQUENCE 419 AA; 45126 MW; BAF29D24BFF81652 CRC64;

Alignment Scores:

Pred. No.: 2.1e-27 Length: 419
 Score: 423.50 Matches: 87
 Percent Similarity: 60.57% Conservative: 62
 Best Local Similarity: 35.37% Mismatches: 96
 Query Match: 12.79% Indels: 1
 DB: 16 Gaps: 1

US-09-868-987-1 (1-1864) x Q8XJ19 (1-419)

QY 65 CTTGATGGCGTATGCTGATGATGACGCTTATATGTCAGACCCCTATTTAAAC 124
 DB 173 LysGlyGlnLysIleAlaIleLysMetAspAsnGlnThrLeuThrAspProValValAsn 192
 QY 125 GTCCCATTTGAAAT---CATGCCAGTCTCAGGGGAAATTTACCAACCGTGAAGTACG 181
 DB 193 AspIleIleSerAsnGlyGlnAlaIleIleSerLysAsnArgSerMetGlnGlnAlaGlu 212
 QY 182 AAATCCGCTCAGATTTTAAATCTGAGCGATGCTTTTGTCCGAGTTCCTCAGTGA 241
 DB 213 LysValSerGlyIleIleAsnAlaGlyAlaLeuProValProValLysAlaValSerVal 232
 QY 242 GAGACGATCTCTTGTATTTGGGAAAAACAATGTACACAGGCAATTTCTCAGCATGC 301
 DB 233 GlnThrValGlyAlaGlnLeuGlyAlaAsnAlaLeuProAsnAlaLeuLysAlaGlyAla 252
 QY 302 TGTGCGTGGCAATGCTTATTTGATGAGCGTATATATATGATTTGGAGCGCTGATC 361
 DB 253 IleGlyValAlaIleIlePheLeuPheMetIleLeuTyrThrValProGlyPheIle 272
 QY 362 GCTTCGGGAGCTGTTCTTGAATCTTTTGTCTTATTCGGGACGCTCACAATATTTGAT 421
 DB 273 AlaCysMetSerLeuSerValTyrIleLeuLeuValLeuTyrIlePheAlaLeuValGly 292
 QY 422 GCGGCACTCACTTGTGAGGATCTGCGGATGTTGTTGCTATGGGATGGCGGTAGAT 481
 DB 293 ValThrLeuThrLeuProGlyIleAlaAlaPheLeuLeuThrValGlyMetAlaValAsp 312

QY 482 GCRAATCTCTGTTATTCGGAAGATCCGACAGCAATTTTATTGTCTCAAGTCTTAAA 541
 Db 313 AlaasnValLeuIlePheGluArgIleGlyLeuGluArgSerIleThr 332
 QY 542 AAATCTGTAGAAAAGGATATACCAAGGCTTTTGGAGCCATTTTGTGATCTCAAGTCTGACT 601
 Db 333 SerAlaMetIleGlyPheSerAsnAlaLeuArgSerIleMetAspSerAsnIleThr 352
 QY 602 ACAGTATTTGGCTCAGCACTCTTTCTTCTTCATGATACAGGCGCTTTTAAAGGTTTGTCT 661
 Db 353 ThrLeuIleAlaGlyLeuValLeuThrPhePheGlySerGlyProValLysGlyPheAla 372
 QY 662 TTGCATCTGTTAGCAATTTCTCTCAATGTTTACGGCTCTTTTCATGACCTAAATTT 721
 Db 373 LeuThrLeuLeuIleGlyIleValIleSerMetPheThrAlaIleIleMetThrArgPhe 392
 QY 722 TTCCTTCATCTGCTGGATGAATAAGACCCCAACATACACAGTTGTCATATGATGAATAAGTTC 781
 Db 393 PheMetAsnLeuGlyPheAsnMetGlyIleLeuAsnLysProSerMetPheGlyArgIle 412
 QY 782 GTGGGGATAAGCATGAT 799
 Db 413 LysGlyGlyArgHisAsn 418

RESULT 21
 Q9PNK1 PRELIMINARY; PRT; 526 AA.
 AC Q9PNK1;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Protein-export membrane protein.
 GN SEC2 OR CJ1093C.
 OS Campylobacter jejuni.
 OC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;
 OC Campylobacter.
 OC Campylobacter.
 OX NCBI_TaxID=197;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NCTC 11168;
 RX MEDLINE=20150912; PubMed=10688204;
 RA Parkhill J., Wren B.W., Mungall K., Ketley J.M., Churcher C.,
 RA Basham D., Chillingworth T., Davies K.M., Feltwell T., Holtroyd S.,
 RA Jagsels K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W.,
 RA Quail M.A., Rajandream A.G., Rutherford K.M., van Vliet A.H.M.,
 RA Whitehead S., Barrett B.G.;
 RT "The genome sequence of the food-borne pathogen Campylobacter jejuni
 reveals hypervariable sequences.";
 RL Nature 403:665-668(2000).
 DR EMBL; ALJ39077; CAB73348.1; -.
 DR InterPro; IPR001036; Acriflavin res.
 DR InterPro; IPR003335; SecD_SecE.
 DR Pfam; PF02355; SecD_SecE.
 DR PRINTS; PR00702; ACRIFLAVINRP.
 DR TIGRFAMs; TIGR00916; 2A0604s01; 1.
 DR TIGRFAMs; TIGR01129; secD; 1.
 KW Complete proteome.
 SQ SEQUENCE 526 AA; 57251 MW; B2DC392902810CID CRC64;

Alignment Scores:
 Pred. No.: 1.06e-25 Length: 526
 Score: 403.50 Matches: 92
 Percent Similarity: 57.60% Conservativeness: 52
 Best Local Similarity: 36.80% Mismatches: 97
 Query Match: 12.18% Indels: 9
 DB: 16 Gaps: 2

US-09-868-987-1 (1-1864) x Q9PNK1 (1-526)

QY 29 ATCAGCGCACTGCTATGGAATATTTCT-----GCCAAC 64
 Db 263 IleAsnPheThrLeuAsnAlaGluGlySerLysLysPheAlaAspThrGlyAlaAsn 282

QY 65 CGTGGATGGCGTATGGCTGTAGTATTGACGGTTATATGTGTCAGCAGCCCTATTTTAAAC 124
 Db 283 ValGlyLysArgAlaIleValLeuAspAsnLysValTyrSerAlaProSerIleAsn 302
 QY 125 GTCCCATG--AAAATCATCCAGTCTCAGGGAAATTTACCCACCGCTGAAGTCAGC 181
 Db 303 GluArgIleGlyGlySerGlyGlnIleSerGlyAlaPheThrGlnGluGluAlaArg 322
 QY 182 AAATCGCTCAGATTTAAATCTGGAGCGATGTCTTTTGTCTCCGAGGTTCTCAGTGAA 241
 Db 323 AspValAlaValAlaLeuArgSerGlyAlaLeuLeuAlaProValLysLeuLeuGluGln 342
 QY 242 GAGACGATCTCTTCTGATCTCTGGAAAAACAATGTACACAGCATTTATCTCAGCATGC 301
 Db 343 ArgSerIleGlyProSerLeuGlyAlaAspSerIleLysMetSerMetIleAlaLeu 362
 QY 302 TGTGGCTTGGCAATGCTTATTGTTGATCAGCGTATATTATATAGATTGGAGGGCTCATC 361
 Db 363 GlyAlaSerValPheIleValValPheMetMetTyrTyrGlyValAlaGlyIlePhe 382
 QY 362 GCTTGGGAGCTGTCTTCTGAATCTTTTCTTATCTGGGCGAGCTCTACAGTATTTGGAT 421
 Db 383 AlaAsnIleAlaMetLeuValAsnValValValAlaValAlaMetAlaMetPheGly 402
 QY 422 GCGCCACTCCTTGTGTCAGACTCGCTGGGATGTTCTTCTGATGGGATGGCCGTAGAT 481
 Db 403 AlaThrLeuThrLeuProGlyMetAlaGlyLeuValLeuThrValGlyMetAlaValAsp 422
 QY 482 GCAATGTTCTTCTGATTCGAAAGTCCGAGAGGAAATTTTATTGTCTCAAGTCTTAAA 541
 Db 423 AlaAsnValIleIleAsnGluArgIleArgIleLeuLeuArgAspGlyValAsnIleArg 442
 QY 542 AAATCTGTAGAAAAGGATATACCAAGGCTTTTGGAGCCATTTTGTGATCTCAAGTCTGACT 601
 Db 443 ValSerIleGluGlnGlyTyrLysAsnAlaMetSerAlaIleIleAspSerAsnIleThr 462
 QY 602 ACAGTATTTGGCTCAGCACTCTTTTCTTCTGATACAGGCGCTTATTAAGGTTTGTCT 661
 Db 463 SerLeuValThrSerValAlaLeuTyrAlaTyrGlyThrGlyAlaValLysGlyPheAla 482
 QY 662 TTGCATCTGTTAGCAATTTCTCTCAATGTTTACGGCTCTTTTCATGACCTAAATTT 721
 Db 483 ValThrLeuGlyIleGlyIleValValSerMetIleThrAlaIleLeuLeuGlyThrHisGly 502
 QY 722 TTCCTTCATCTGCTGGATGAATAAGACCCAA 751
 Db 503 MetPheAspTyrPheMetGlnArgIleGlu 512

RESULT 22
 Q8RLP5 PRELIMINARY; PRT; 518 AA.
 AC Q8RLP5;
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Export membrane protein SecD.
 OS Rickettsia typhi.
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
 OC Rickettsiaceae; Rickettsiae; Rickettsia.
 OX NCBI_TaxID=785;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA On'gele B.A., Radulovic S., Azad A.P.;
 RT "Characterization of the Sec operon of the typhus group Rickettsia.";
 RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AY078352; AAL82790.1; -.
 SQ SEQUENCE 518 AA; 57125 MW; 6E9E2F2FB717201F CRC64;

Alignment Scores:
 Pred. No.: 1.91e-25 Length: 518
 Score: 400.50 Matches: 87
 Percent Similarity: 58.40% Conservativeness: 52
 Best Local Similarity: 36.55% Mismatches: 98

Query Match: 12.09% Indels: 1
DB: 2 Gaps: 1
US-09-868-987-1 (1-1864) x Q9CP37 (1-518)

QY 62 AACGGTGGATGGCGATGCTGTGATGATGACGGTATATGATGACGACCGCTATTTTA 121
DB 281 AsnThrGlyLysArgLeuAlaIleValLeuAspAsnLysLeuSerAlaProMetIle 300
QY 122 AAC---GTCCATTGAAAAATCATGCCAGTCTCTCGAGGAAATTTACCCACCGTGAATG 178
DB 301 AsnGlyAlaIleIleGlyValAspGlyLeileThrGlyAsnPhenThrIleGlySerAla 320
QY 179 AGCAAACTGCGCTCAGATTAAATCTGAGCGATGCTTTTGTCCGAGAGTCTCATGT 238
DB 321 AsnGlyLeuAlaIleLeuLeuArgValGlySerLeuProThrProLeuLysIleIleGlu 340
QY 239 GAAGAGACGATCTCTTCTGATCTTGGGAAAAACAATGATACACAGGATATCTCAGCA 298
DB 341 GluArgSerIleGlyProAsnLeuGlyAlaAspSerIleGlySerGlyLysAlaGly 360
QY 299 TGCTGTGGCTGGCAATGCTTATTTGATGACGCTATATATGATTTGAGGCGCTC 358
DB 361 LeuIleGlyPheThrAlaValCysIlePheMetIleLeuSerIleGlyValIleGlyLeu 380
QY 359 ATCGCTTGGAGAGCTGTTCTTCTGATTTTGTCTTATCTGGGACGCTCTACAGTATTG 418
DB 381 PheAlaSerIleAlaLeuIleLeuAlaLeuLysLeuLysLeuLysLeuLysLeuLysLeu 400
QY 419 GATGGCCATCAGCTTGTGACGATCGCTGGGATGTTCTTGTGATGAGGATGCGCGTA 478
DB 401 GluAlaThrLeuThrLeuProGlyIleAlaGlyIleIleLeuThrMetGlyMetAlaVal 420
QY 479 GATCGAAATGTTCTGTGATTCGAAAGATCCGAGAGATTTTATGTCACAGTCT 538
DB 421 AspAlaAsnValLeuIleIleIleIleIleIleIleIleIleIleIleIleIleIleIleIle 440
QY 539 AAAAAATCTGAGAAAAAGATATACCAAGCTTTTGGAGCCATTGATTTCACTTG 598
DB 441 LeuIleAlaIleArgThrGlyPheGlySerAlaPheAlaThrIleIleAspSerAsnLeu 460
QY 599 ACTACAGATGGCTGCTGACACTTCTTCTTCTTCTGATACAGGCGCTATTAAGGCTT 658
DB 461 ThrThrLeuIleValAlaPheAlaLeuIlePheGlyValGlyAlaIleLysGlyPhe 480
QY 659 GCTTGAATGATGATTTGATTTCTTCTTCAATGTTAGCGCTTTTCAATGACTAAA 718
DB 481 AlaValAlaLeuThrIleGlyIleIleSerSerMetPheSerAlaIleIleIleThrLys 500
QY 719 TTTTCTTCATGCTGTGATGAATTAAGCCAACTACACAGTTGCATATGATG 772
DB 501 LeuLeuIleAspIleThrValIleIleIleIleIleIleIleIleIleIleIleIleIle 518
RESULT 23
Q9CP37 PRELIMINARY; PRT; 616 AA.
AC Q9CP37;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE SecD.
GN SEC D OR PM0227.
OS Pasteurella multocida.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Pasteurella
NCBI_TaxID=747;
OX NCBI
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PM70;
RX MEDLINE=2145866; PubMed=11248100;
RA May B.J., Zhang Q., Li L.L., Paustian M.L., Whitlam T.S., Kapur V.,
RT "complete genomic sequence of Pasteurella multocida Fm70";
RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).

DR EMBL: AB006057; AK02311.1; -
DR InterPro: IPR001036; Acrlvln_res.
DR InterPro: IPR003335; SecD_SecF.
DR Pfam: PF02355; SecD_SecF.1.
DR PRINTS: PR00702; ACRIPLAVINRP.
DR TIGRPFAMs: TIGR00916; 2A0604601.1.
DR TIGRPFAMs: TIGR01129; secD.1.
KM Complete proteome.
SQ SEQUENCE 616 AA; 67401 MM; 98B98497DFB8403E CRC64;
Alignment Scores:
Pred. No.: 2,36e-25 Length: 616
Score: 399.50 Matches: 83
Percent Similarity: 66.34% Conservative: 53
Best Local Similarity: 40.49% Mismatches: 58
Query Match: 12.06% Indels: 11
DB: 16 Gaps: 2
US-09-868-987-1 (1-1864) x Q9CP37 (1-616)

QY 137 AATGATCCAGTGTCTCAGGAAATTT-----ACC 166
DB 395 AsnValAlaThrIleGlyArgPheGlySerGlnPheGlnIleThrGlyIleAspSer 414
QY 167 CACCGTAAGTGACCAACTCGCTCAGATTAAATCTGAGCGATGCTTTTGTCC 226
DB 415 ProAlaGlnAlaLeuAsnLeuAlaValLeuLeuArgSerGlyAlaLeuIleAlaProIle 434
QY 227 GAGGTCTCAGTAAGAGACGATCTTCTGATCTTGGGAAAAACAATGATACACAGGC 286
DB 435 GlnIleValGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 454
QY 287 ATTATCTCAGACAGCTGCGCTTGGCAATCTTATGTTGATGACGCTATATTAAGA 346
DB 455 LeuGlnAlaSerPheThrGlyLeuLeuAlaValIleIlePheMetThrLeuIleIleArg 474
QY 347 TTTGAGGCGCTCATCGCTTGGGAGCTGTTCTTCTGATCTTTTGTCTTATCTGGCAGCT 406
DB 475 LysPheGlyLeuIleAlaAsnIleAlaLeuIleAlaAsnIleValIleValIleVal 494
QY 407 CTACAGTATTG--GATGGCCACTGACCTTGCAGACCTGCGGATGTTGTTGCT 463
DB 495 MetSerLeuLeuProGlyAlaThrLeuSerMetProGlyIleAlaGlyIleValLeuSer 514
QY 464 ATGGGATGGCGCTGATGATGCAATGTTCTTGTATTCGAAAGATCCAGAGCAATTTTAA 523
DB 515 ValGlyMetSerIleAspAlaAsnValIleIleIlePheGlnArgIleLysGlnGlnIleArg 534
QY 524 TTGCTCAAGCTCTTAAATAATCTGTAGAAAAAGATATACCAAGCTTTTGGAGCAT 583
DB 535 AsnGlyArgProIleGlnGlnAlaIleIleIleGlnGlyIleIleGlnGlyIleIleSerIle 554
QY 584 TTTGATCTTAAGTCTTACAGTATTTGGCTGACGACTCTTTTCTTCTTCTGATACAGG 643
DB 555 PheAspAlaAsnLeuThrIleThrIleLeuThrSerValIleLeuIleValAlaGlyIle 574
QY 644 CTTATTAAGGCTTGTGATGATTTTGAATTTTCTTCAATGTTTACGCT 703
DB 575 ProValGlyGlyPheAlaIleThrLeuAlaLeuGlyValGlyIleSerMetPheThrAla 594
QY 704 CTTTCTGACTAAA 718
DB 595 IleValGlyThrArg 599
RESULT 24
Q92H77 PRELIMINARY; PRT; 518 AA.
AC Q92H77;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Protein-export membrane protein secD.
GN SEC D OR RC0894.


```

Db 293 ThrLeuThrLeuSerGlyIleAlaGlyLeuLeuLeuThrIleGlyMetAlaValAlaSerAla 312
Qy 485 AATGCTCTTGTATTCGAAAATAATCCGAGAGAAATTTTATGTCCTCAAGCTTTAAATAA 544
Db 313 AsnValIleuIlePheGluAlaGlyPheLysGluGluLeuLysIleGlyLysThrValAlaSer 332
Qy 545 TCTGTAGAAAAAGATATACCAAGCTTTTGAGCCATTTTGTGATCTTCACTGATCA 604
Db 333 AlaPheAspAlaGlyPheIleAspAlaMetSerSerIleIleAspSerAsnValThrThr 352
Qy 605 GTATTGGCTCAGACACTCTTTCTTCTTCTAGATACAGGCGCTATTAAGGCTTGTG 664
Db 353 IleIleSerGlyValIleLeuThrAlaLeuGlySerGlyValGlyValGlyPheAlaLeu 372
Qy 665 ACATTGATTTTAGAATTTCTCTTCAATGTTTACGCGCTTTTTCATGACTTAATTTTC 724
Db 373 ThrLeuValIleGlyValIleLeuIleSerMetValThrAlaIleIleValThrArgArgLeu 392
Qy 725 TTCATGCTGTGG-----ATGAATAGACCCCAACAT 754
Db 393 LeuMet---TTPAlaIleAspMetGlyLeuIleAsnLysAlaSerHis 407

RESULT 26
ID 09A6U2 PRELIMINARY; PRT; 532 AA.
AC 09A6U2;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Protein-export membrane protein Secd.
GN CCI1991.
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
OC Caulobacter.
NCBI_Taxid=155892;
RX NCBI [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 19089 / CB15;
RX MEDLINE=21173698; PubMed=11259647;
RA Nieman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA Potocka I., Nelson W.C., Newton A.S., Stephens C., Phadke N.D., Ely B.,
RA Deboy R.T., Dodson R.U., Durkin A.S., Gwinn M.L., Hatt D.H., Berry K.,
RA Kolonay J.F., Smit J., Craven M.B., Knouri H., Shetty J., Berry K.,
RA Utechtack T., Tran K., Wolf A., Vamathevan J., Ermoleva M., White O.,
RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.,
RT "Complete genome sequence of Caulobacter crescentus.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
DR EMBL; AB005872; AAK23966.1; -
DR TIGR; CCI1991; -
DR InterPro; IPR003335; Secd_SecF.
DR Pfam; PF02355; Secd_SecF; 1.
DR TIGRFAMs; TIGR00916f_2A0604s01; 1.
DR TIGRFAMs; TIGR01129; secD; 1.
KW Complete proteome.
SQ SEQUENCE 532 AA; 56792 MW; 80A85984798F64C7 CRC64;

Alignment Scores:
Pred. No.: 1.22e-24 Length: 532
Score: 291.00 Matches: 85
Percent Similarity: 60.34% Conservative: 58
Best Local Similarity: 35.86% Mismatches: 90
Query Match: 11.81% Indels: 4
Gaps: 3

US-09-868-987-1 (1-1864) x 09A6U2 (1-532)
Qy 62 AACCGTGAATGGCGTATGCTGATGATGAAGTATATGCTGAGAGCCCTATTTTA 121
Db 295 AsnValIleuLysArgPheAlaIleValIleuAspGlyArgIleIleSerAlaIleProThrIle 314
Qy 122 AAC--GTCCATTGAAAAAATCATGCCAGCTGTCTAGAGGAAATTTTACCACCGTGAAGTG 178

```

```

Db 315 AsnGlyAlaIleLeuGlyGlySerGlyIleIleIleThrGlySerPheThrAlaGluSerAla 334
Qy 179 AGCAACTCGCTCAGATTTTAAATCTGAGCCGATGCTTTGTTCCGAGCTTCTAGT 238
Db 335 SerAspLeuIleLeuLeuLeuLeuArgSerGlyAlaLeuProAlaProLeuLysValGluGln 354
Qy 239 GAAGAGCGAATCTTCTGATCTTGAATCTTGAATCTTGAATCTTGAATCTTGAATCTTGA 298
Db 355 GluAsnThrValGlyAlaGluLeuGlyAlaAspAlaValAlaGlyAlaIleSerThr 374
Qy 299 TGCTGTGGCTTGGCAATGCTTATTTGTTGATGAGCCGATATTTATTTATTTATTTATTT 352
Db 375 LeuValAlaPheIleThrIleValIleValPheMetIleLeuSerThrGlyLeuLeuPheGly 394
Qy 353 GGGCTATCGCTTGGAGAGCTGTTCTTCTGATCTTGTGCTTATCTGGGAGCTCTACAG 412
Db 395 Gly---IleSerValIleAlaLeuIleIleAsnGlyMetLeuIleValAlaAlaMetSer 413
Qy 413 TATTGGATGGCGACATCCACTTGTCTGAGAGCTGCTGGGATGTTCTTGTGATGGGATG 472
Db 414 LeuThrGlnAlaThrLeuThrLeuProGlyIleAlaGlyLeuIleLeuThrLeuAlaVal 433
Qy 473 GCCGTAGATGCAATGTTCTTGTATTCGAAAGATCCGAGAGAAATTTTATTTATTTCTCAA 532
Db 434 AlaValAspAlaAsnValLeuIleIleIleIleIleIleIleIleIleIleIleIleIle 453
Qy 533 AGCTTTAAAAATCTGAGAAAAAGATATACCAAGCTTTTGAGCCATTTTGAATCT 592
Db 454 SerProIleLeuAlaAlaAspAlaGlyPheSerArgAlaMetThrThrIleIleAspAla 473
Qy 593 AACTGACTACGATTTGGCTGACGACTCTTCTTCTCTAGATACAGGCGCTATTA 652
Db 474 AsnValThrThrLeuValAlaAlaGlyIleMetPheAlaPheGlyAlaGlyProValArg 493
Qy 653 GGGTTTCTTGAATTTTGAATTTTGAATTTTCTTCAATGTTTACGCTTTTTCATG 712
Db 494 GlyPheAlaIleThrLeuSerIleGlyValPheHisSerValPheThrAlaValLeuVal 513
Qy 713 ACTAAATTTTCTTCATGCTGTGATGATATACCAACCAATACACAGCTTG 763
Db 514 SerGluLeuLeuIleGlyTyrTrpPheArgAlaAlaArgProLysLysLeu 530

RESULT 27
ID 09PNK2 PRELIMINARY; PRT; 323 AA.
AC 09PNK2;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Protein-export membrane protein.
GN SecF OR CUI092C.
OS Campylobacter jejuni.
OC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;
OC Campylobacter.
NCBI_Taxid=197;
RX NCBI [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCTC 11168;
RX MEDLINE=20150912; PubMed=10688204;
RA Parkhill J., Wren B.W., Mungall K., Kerley J.M., Churcher C.,
RA Basham D., Chillingworth T., Davies R.M., Feltham J., Holtroyd S.,
RA Jørgensen K., Kariyasek A.V., Moule S., Pallen M.J., Penn C.W.,
RA Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M.,
RA Whitehead S., Barrett B.G.;
RT "The genome sequence of the food-borne pathogen Campylobacter jejuni
reveals hypervariable sequences.";
RL Nature 403:665-668(2000).
DR EMBL; AL139077; CA873347.1; -
DR InterPro; IPR003335; Secd_SecF.
DR Pfam; PF02355; Secd_SecF; 1.
DR TIGRFAMs; TIGR00916f_2A0604s01; 1.
DR TIGRFAMs; TIGR00966; 3a0501s07; 1.

```



```
KW Complete proteome.
SQ SEQUENCE 323 AA; 36112 MW; 23589FD95B5BFD02 CRC64;

Alignment Scores:
Pred. No.: 1,41e-24 Length: 323
Score: 390.00 Matches: 98
Percent Similarity: 53.09% Conservative: 74
Best Local Similarity: 30.25% Mismatches: 96
Query Match: 11.78% Indels: 56
DB: 16 Gaps: 9

US-09-868-987-1 (1-1864) x Q9PNK2 (1-323)
QY 794 CATGATTTCTTGAGAGATGCAAAAACCTTTGGCTGTTCTT----- 835
Db 10 TyrAspPheMetArg-----MetArgPheAlaIleSerLeuSerPheIleLeuPhe 27
QY 836 ---GGAAGTGTTCCTTTAGTTGGCTTCTCGGTTGGAGCTGGAATTCGTT 892
Db 28 PheGlySerIleThrLeuLeu-----TrpAspArgGly 38
QY 893 TTG-----GGAATGGATTTTAAAGGAGGATGCTTTTACCTTTAATCCAAAAGAGCAT 946
Db 39 LeuGlnThrGlyIleAspPheSerGlyThrLeuValGlnLeu-----LysThrGlu 56
QY 947 GGCATCAGCGATGTGCTCAAAATCGTGGCAAAAGTTGTGCATAACTACAGGAAGCTGGT 1006
Db 57 AsnAlaIleProIleThrGlnIleArg-----GluIleLeuGluAsnGln 71
QY 1007 CTTCTCTAGAGACTCCGTTATCAACATTTGGAATCTCAGAAAGATCAAAATCTAT 1066
Db 72 GlyThrPheGlnAsnLeuSerValThrGluPheGlySerAsnGluValThrIleArg 91
QY 1067 TTT-----AGTGATAAAGCTTTAAGCTATATAAGCAGATACAGGCTCTCTCTA 1117
Db 92 PheLeuGlySerAsnAspAsnValSerAsnAspIleGlyGluIleSerThrLeuLeu 111
QY 1118 AAATTAACGATCATGAGCTGGCTGTTATGTGGGATGTGTGTGAGAAACAGCGCTAGATT 1177
Db 112 LysAspThr----- 114
QY 1178 CTCTACGGAACCTTAACAGAAACGCAAAATTTGGTCAAAAGGTAAAGCAGCAAACTATCG 1237
Db 115 -----GlyLysPheGluValArgAlaAspValValGlyProLysValGly 130
QY 1238 AAGAAATGCTTATCAGCGCACCATCGGCTTTTAGGAGCTTTGGCAATCATCTTGCTC 1297
Db 131 AspGluLeuArgAsnLysGlyLeuMetAlaIleAlaValSerLeuIleAlaIleLeu 150
QY 1298 TATGTGAGTTTGGCTTTGAATGGCAATATGCTTTAGTGGCTGATGCGCTTTAATTCAT 1357
Db 151 TyrIleAlaLeuArgPheGluThrArgPheAlaLeuAlaIleIleSerGluIleHis 170
QY 1358 GACCTTTGGCTACTGTGCACTGTTGTTATAGCACATTTCTTTTGAAGAAATTCAA 1417
Db 171 AspValIleThrLeuGlyAlaIle-----SerLeuPheLysIleAsp 185
QY 1418 ATAGATTGCAAGCATTTGGTCTTTAATGACTGTATTGGGTATTCATTAAACAATACT 1477
Db 186 ValAsnLeuAspThrLeuAlaAlaValLeuThrValLeuGlyThrSerLeuAsnAspThr 205
QY 1478 TTGATCATTTTGTGATTCGTGATGATCGCAAGGCAACCTGTTTACCCCTATGCA 1537
Db 206 IleIleIlePheAspArgIleArgGluGlyIleLysThrSerLysLysThrGluLeuAla 225
QY 1538 GTTTTAGTTAATGATGCCCTTCAAAAGAGCTTACGCCACGGTAATGAACAGCTACA 1597
Db 226 ProIleIleAsnGluSerValSerAlaThrLeuSerArgThrValLeuThrSerGlyLeu 245
QY 1598 ACTCTATCAGTTTGTAAATCTTTGTTTATAGCGGCTCTCTGCTTTTAAATTTGCA 1657
Db 246 ThrLeuAlaThrValIleLeuThrPhePheGlyGlyGluMetIleGlnGlyPheSer 265

QY 1658 TTTATTATAGACCATAGGATTCCTTTAGGAACCTTTATGCTCTCTTATATATGACCACT 1717
Db 266 LeuAlaLeuIleValGlyIleIleAlaGlyThrLeuSerSerIlePheValAlaSerPro 285
QY 1718 CTGTTGTTGTTT 1729
Db 286 ThrLeuLeuThr 289

RESULT 28
Q8ZC36 PRELIMINARY; PRT; 322 AA.
ID Q8ZC36;
AC Q8ZC36;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-JUN-2002 (TrEMBLrel. 20, Last sequence update)
DE Protein-export membrane protein SecF.
GN SEC F OR YPOJ188.
OS Yersinia pestis.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Yersinia.
ON NCBI_TaxID=632;
RX [1]
SEQUENCE FROM N.A.
STRAIN=CO-92 / BIOVAR ORIENTALIS;
MEDLINE=21470413; PubMed=11586360;
RA Parkhill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G.,
RA Prentice M.B., Sebahia M., James K.D., Churcher C., Mungall K.L.,
RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdono-Tarraga A.M.,
RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
RA Feitwell T., Hamlin N., Holroyd S., Jagels K., Karlyshev A.V.,
RA Leather S., Moulé S., Oyston P.C.F., Quail M., Rutherford K.,
RA Simmonds M., Skellerton J., Stevens K., Whitehead S., Barrell B.G.;
RT "Genome sequence of Yersinia pestis, the causative agent of plague.";
RL Nature 413:523-527(2001).
DR EMBL; AJ414155; CAC92423.1; -
DR InterPro; IPR003335; SecD_SecF.
DR Pfam; PF02355; SecD_SecF; 1.
DR TIGRFAMs; TIGR00916; 2A0604801; 1.
DR TIGRFAMs; TIGR00966; 3a0501s07; 1.
KW Complete proteome.
SQ SEQUENCE 322 AA; 35178 MW; 2029AEB45F281ADC CRC64;

Alignment Scores:
Pred. No.: 4.13e-24 Length: 322
Score: 384.50 Matches: 104
Percent Similarity: 52.88% Conservative: 61
Best Local Similarity: 33.33% Mismatches: 114
Query Match: 11.61% Indels: 33
DB: 16 Gaps: 10

US-09-868-987-1 (1-1864) x Q8ZC36 (1-322)
QY 794 CATGATTTCTTGAGAGATGCAAAAACCTTTGGCTGTTCTTGGAACTGTTTTCTTTA 853
Db 17 TyrAspPheMetArgTrpAspTyrValAlaPheGlyValSerLeuLeuLeuVala 36
QY 854 GGTTCGCTGCTCTCGGTTTGGAGCTGCAATTCCTGTTTGGGAATGATTTAAAGGA 913
Db 37 SerIleValMetSerThrLysGlyPheAsn-----TrpGlyLeuAspPheThrGly 54
QY 914 GGGTATGCTTTACCTTTAATCCAAAAGAGCATGGCATCAGCGATGTTCCTCAATGCGT 973
Db 55 GlyThrValIleGluIleAsnLeuGluAsnPro-----AlaAspLeuAspGlnLeuArg 72
QY 974 GGCNAATGTTGCATAAACTACAGGAAGCTGTTCTTCTTAGAGACTTCCTCGATTCAA 1033
Db 73 -----AspThrLeuGlnAsnAlaGlyPheGluSerPro-----IleLeuGln 86
QY 1034 ACATTTGATCTTCAGAAAGATCAAAATCTATTTTAGTGATAAAGCTTTAAGCTATACT 1093
Db 87 AsnPheGlySerArgAspValMetValArgMetProAlaThrGlyThrAlaGly 106
QY 1094 AAGCAGATACGAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1153
```

```

Db 107 GlnGluLeuGlyAenLysIleIleSerValIleAsnGluSer----- 120
Qy 1154 GTTGTGAGAAACAGGCGCTAGATTCTTACGAAACTCTAAACGAAACGAAATTTGG 1213
Db 121 ValAspLysAsnAla-----SerValLysArgIleGluPhe--- 132
Qy 1214 TCAGAGTAACGACCAACTCTCGAAGAAATGCGTTAGCGGAGGACCATGGGCTTTA 1273
Db 133 -----ValGlyProSerValGlySerAspLeuAlaGlnAlaGlyAlaLeuAlaLeu 150
Qy 1274 GGAGCTTTGGCAATCATCTGCTATGAGTTTGCGCTTTGATGGCAATATGCTTTC 1333
Db 151 ValAlaLeuLeuSerIleLeuValIleValGlyPheArgPheGluIlePargLeuAlaLeu 170
Qy 1334 AGTCCCGTATGCGCTTAATTCATGACCTTTGGCTACCTTGACGCTTGTATAGCA 1393
Db 171 GlyAlaValIleSerLeuAlaHisAspValValIleThrMetGlyIleLeuSerLeuPhe 190
Qy 1394 CATTTCTTTTGAAGAAATTCMAATAGATTGGCAAGCATTTGGCTTTAATGACTGA 1453
Db 191 His-----IleGlnIleAspLeuThrIleIleAlaSerLeuMetSerVal 205
Qy 1454 TTGGGGTATTCATTAACAATACATCTTATGATTTTGTATCGTATTCGGAAGAT---CGC 1510
Db 206 IleGlyIleSerLeuAspSerIleValAlaSerAspAlaIleArgGluAsnPheArg 225
Qy 1511 CAAACGAACTGTTTACCCCTATGATGCTTTTAAATGATGATGCTTCAAAAGACGTTT 1570
Db 226 LysIleArgArgGlyThrProTyrGluIleMet---AsnValSerLeuThrGlnThrLeu 244
Qy 1571 AGCCGCAAGGTAATGACAAACGCTTACATGCTTATGCTTTGTAAGCTTTTGTATTA 1630
Db 245 SerArgThrIleMetThrSerAlaThrThrLeuMetValValLeuMetLeuPheIlePhe 264
Qy 1631 GGGCGCTCCTGCTTAAATTTTGGCATTATTATGACATGAGGATCTTCTAGAACT 1690
Db 265 GlyAlaValaMetLeuGlnGlyPheSerLeuThrMetLeuIleGlyValThrIleGlyThr 284
Qy 1691 TTATGCTCTCTTATATATGACCAACTCTGTTGTTG 1726
Db 285 ValSerSerIleTyrValAlaSerAlaLeuAlaLeu 296

RESULT 29
Q82C35 PRELIMINARY; PRT; 615 AA.
AC Q82C35;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Protein-export membrane protein SecD.
OS SecD OR YP03189.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Yersinia.
ON NCBI_Ftaxid=632;
RX MEDLINE=21470413; PubMed=11586360;
RA Parkhill J., Wren B.W., Thomson N.R., Titchball R.W., Holden M.T.G.,
RA Prentice M.B., Sebahia Y., James K.D., Churcher C., Mungall K.L.,
RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdeno-Tarraga A.M.,
RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
RA Feltwell T., Hamlin N., Holroyd S., Jagsels K., Karlyshev A.V.,
RA Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford K.,
RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrall B.G.;
RT "Genome sequence of Yersinia pestis, the causative agent of plague.";
RL Nature 413:523-527 (2001).
DR EMBL; AJ141455; CAC92424.1; -
DR InterPro: IPR003335; SecD SecP.
DR InterPro: IPR005008; SigPase.
DR Pfam; PF02355; SecD_SecP; 1.

```

```

DR TIGRPFAM; TIGR00916; 2A0604801; 1.
DR TIGRPFAM; TIGR01129; secD; 1. UNKNOWN_1.
DR PROSITE; PS00761; SPASE_1_3; UNKNOWN_1.
KW Complete proteome.
SQ SEQUENCE 615 AA; 66833 MW; 3F48E8014975CPCF CRC64;

Alignment Scores:
Pred. No.: 5,35e-24 Length: 615
Score: 383.50 Matches: 86
Percent Similarity: 62.83% Conservative: 56
Best Local Similarity: 38.05% Mismatches: 81
Query Match: 11.58% Indels: 3
DB: 16 Gaps: 2

US-09-868-987-1 (1-1864) x Q82C35 (1-615)
Qy 44 AATGACAAATATTCGAAACCGTGAATGGCGTATGCTAGTGAATGATATG 103
Db 375 SerGlyLysLysAspAlaAsn-----GlyArgSerIleLeuValLysGlnGluVal 392
Qy 104 GTCAGACGCCCTATTTAAACGTCCTCCATTTGAAATATCATGCCAGTCTCAGGAAATTT 163
Db 393 IleAsnValAlaThrIleGlnSerArgLeuGlyAsnSerPheArgIleThrGlyIleAsp 412
Qy 164 ACCACCGTGAAGTGAACCAACTCGCCTCAGATTTAATCTGAGGAGATGCTTTGTT 223
Db 413 AsnProAlaGluAlaArgGlnLeuSerLeuLeuMetArgAlaGlyAlaLeuIleAlaPro 432
Qy 224 CCCGAGTTCTCAGTGAAGAGACGATCTCTTGCATCTTGGGAAAAACAATGTACAA 283
Db 433 IleGlnIleValGlnGluArgThrIleGlyProThrLeuGlySerGlnAlaGln 452
Qy 284 GGCATTATCTCAGACGCTGTGCTTGGCAATGCTTATTTGTAAGCCTATATAT 343
Db 453 GlyLeuGluAlaCysLeuThrPglyLeuAlaValSerIleLeuPheMetValValIleTyr 472
Qy 344 AGATTGAGAGCGTCATCGCTCGGAGCTGTCTTCTGTAATCTTTGCTATCTATCTGGCA 403
Db 473 ArgLysPheGlyValIleAlaSerThrAlaLeuMetAlaAsnLeuValIleValGly 492
Qy 404 GCTCTACAGTATTTG--GATGGCCCACTCACCTTGTACAGACTCGCTGGATTTGTTT 460
Db 493 ValMetSerLeuLeuProGlyAlaThrLeuThrMetProGlyIleAlaGlyIleValLeu 512
Qy 461 GCTATGGGATGGCCGTATGCAATGCAATGCTTCTTATTCGAAAGATCCAGAGGAATTT 520
Db 513 ThrLeuAlaValAlaValaAspAlaAsnValIleLeuIleAsnGluArgIleLysGluGluTyr 532
Qy 521 TTATTTGCTCAAGTCTTAAAAATCGTAGAAAGAGATATCAAGGCTTTTGGAGCC 580
Db 533 ArgAsnGlyArgThrIleGlnGlnAlaIleHisGluGlyTyrLysGlyAlaPheSerSer 552
Qy 581 ATTTTATTCATTCATGCTACAGATATGCGCTTACAGCACTTTTCTTCTAGATACA 640
Db 553 IleValaAspAlaAsnIleThrThrLeuIleThrAlaIleIleLeuTyrAlaValaGlyThr 572
Qy 641 GGGCTATTAAAGGTTTGCTTGCATTTAGCAATTTTCTTCTTAAATGTTTACG 700
Db 573 GlySerIleLysGlyPheAlaIleThrThrAlaIleGlyAlaValThrSerMetPheThr 592
Qy 701 GCTCTTTTCACTATAA 718
Db 593 AlaIleValaGlyThrArg 598

RESULT 30
Q8RAM4 PRELIMINARY; PRT; 410 AA.
AC Q8RAM4;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DE Preprotein translocase subunit SecD.
GN SEC D OR TTE1188.

```

OS *Thermoanaerobacter tengcongensis*.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Thermoanaerobacteriales; Thermoanaerobacteriaceae; Thermoanaerobacter.
NCBI_TaxID=119072;
RN [1]
RX SEQUENCE FROM N.A.
RP STRAIN=MB4T / JCM11007;
RC MEDLINES=21992816; PubMed=11997336;
RX Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,
RA Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,
RA Tan H., Chen R., Wang J., Yu J., Yang H.;
RT "A complete sequence of *T. tengcongensis* genome.";
RL Genome Res. 12:689-700(2002).
RW EMBL; A5013081; AAM24419.1; -.
DK Complete genome.
SQ SEQUENCE 410 AA; 43931 MW; 877F798E3208F13E CRC64;

Alignment Scores:
Pred. No.: 7,598-24 Length: 410
Score: 381.50 Matches: 73
Percent Similarity: 63.98% Conservative: 82
Best Local Similarity: 30.93% Mismatches: 82
Query Match: 11.52% Indels: 3
DB: 16 Gaps: 2

US-09-868-987-1 (1-1864) x QBRAM4 (1-410)

QY	68	GGATGGCGTATGCTAGTGGATTGACGGTATATATGGTGCAGCAGCCCTATTTTA---AAC	124
DB	173	GlyGlnProIleAlaIleLeuLeuAspAspIysValIleSerAlaProValValArgAsp	192
QY	125	GTCCCAATGAAAATCATCCAGTGCTCCAGGGAAATTTACCCACCGTGAAGTCAGCAAA	184
DB	193	ValIleThrThrGlyAsnAlaValIleThrGlyLeuLysAspPheGlnGluAlaSerGlu	212
QY	185	CTGCCTCAGATTTAAATCTGGAGCGATGCTCTTTGTTCGCCAGGTTCTCAGTGAAGAG	244
DB	213	LeuAlaThrLeuIleArgAlaGlySerLeuProValThrLeuLysProIleAlaTyrSer	232
QY	245	ACGATCTCTTCATCTCGGAAAAACAATGTACACAAGGCATTATCTCAGCATGCTGT	304
DB	233	SerValGlyAlaThrLeuGlyProSerAlaLeuLysAlaSerValGluAlaGlyIleTyr	252
QY	305	GGCTTGGCAATGCTTATGTTTGATGACGCGTATATATAGATTTTGGAGGCGTCATCGCT	364
DB	253	GlyThrLeuLeuValMetLeuPheMetMetAlaPheTyrArgLeuProGlyPheIleAla	272
QY	365	TCGGGAGCTGTTCTTCGAAATCTTTTGGCTTATCTGGGCGAGCTCTACAGTATTGGATGCG	424
DB	273	AspLeuAlaLeuLeuIleTyrIleLeuIleAsnPheIleIleTyrAlaIlePheAsnVal	292
QY	425	CCACTCACCTTGTTCAGACTCGCTGGGATGTTGTTCTGCTATGGGATGCGCGTAGTCGA	484
DB	293	ThrLeuAspLeuProGlyIleAlaGlyPheLeuLeuSerIleGlyMetAlaValAspThr	312
QY	485	AATGTTCTTCTGATTCGAAAGAATCCGAGAGGAATTTTATTTGCTCTCAAGTCTTAAAAA	544
DB	313	AsnIleLeuIlePheGluArgPheLysGluGluLeuTrpAlaGlyLysThrIleArgPro	332
QY	545	TCGTGTAGAAAAGGATATACCAAGGCTTTTGGAGCCATTTTGGATTCTTAACCTGACTACA	604
DB	333	AlaLeuAspAlaGlyPheAlaLysAlaMetArgAlaValIleAspAlaAsnValThrThr	352
QY	605	GTATTGGCCTCAGCACTCTTTTCTCTCTAGATACAGGCCTATTAAAGGTTTTCITTTG	664
DB	353	IleIleGlyAlaIleLeuPheTyrLeuGlySerGlyAsnValLysGlyPheAlaLeu	372
QY	665	ACATTGATTTTAGGAATTTCTCTCAATGTTTACGGCTCTTTTTCATGACTAAATTTTC	724
DB	373	ThrLeuLeuIleGlyValSerSerLeuPheThrAlaIleThrValThrArgPheLeu	392
QY	725	TTC-----ATGCTGTGGATGAATAAGACCCCAACATACACATGCTGTCAT	766

```

QY 365 TCGGAGCGTCTTCTTGATCTTTGCTAFTCTGGGACGCTTACAGATTG---GAT 421
Db 482 Asn1leAlaLeuMetAlaAsnleuValleuIleIleGlyValMeSerMetIleProGly 501
QY 422 GCGGCACCTACCTTGTTCAGACGACGCTGGATTTGTTCTTGGTGGGANTGGCGTAGAT 481
Db 502 AlaIleMetIleProGlyIleIleAlaGlyIleValleuIleuIleValGlyMetAlaValasp 521
QY 482 GCAATGTTCTTGTATTCGAAAGATCCGAGAGCAATTTTATTTCTTCAAAAGTCTTAA 541
Db 522 AlaAsnValleuIlePheGlyIleArgIleArgIleuIleuArgIleuIleuAsnProGln 541
QY 542 AAATCTGTAGAAAAGATATACCAAGCTTTTGAGCGCATTTTGAATTAACCTGACT 601
Db 542 GlnAlaIleHisGlnGlyIleAlaAsnAlaPheSerThrIleAlaAspAlaAsnIleThr 561
QY 602 ACAGTATTCGCTCAGACGACCTCTTCTTCTTCTAGATACAGGCGCTATTAAGGCTTTCCT 661
Db 562 ThrIleuIleThrAlaIleIleuPheAlaValaGlyThrGlyAlaIleIleGlyPheAla 581
QY 662 TTGACATTTGATTTAGAAATTTCTCTTCAATGTTTACGGCTCTTTTCATGACTTAA 718
Db 582 ValThrIleuSerIleGlyIleuThrIleuSerMetPheThrAlaIleValaGlyThrArg 600

RESULT 32
Q8XVW2 PRELIMINARY; PRT; 626 AA.
ID Q8XVW2
AC Q8XVW2;
DT 01-MAR-2002 (TReMBLrel. 20, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Probable export membrane protein SEC2 transmembrane.
GN SEC2 OR RSC2715 OR R50081.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
OC Ralstonia.
NCBI_TaxID=305;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GM11000;
RX MEDLINE=21681879; PubMed=11823852;
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA Ariat M., Billault A., Brottier P., Camus J.C., Cattolico L.,
RA Chandler M., Choisme N., Claudel-Renard C., Cunne S., Demange N.,
RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schlex T.,
RA Siguier P., Thebaud P., Whalen M., Wincker P., Levy M.,
RA Weissenbach J., Boucher C.A.;
RT "Genome sequence of the plant pathogen Ralstonia solanacearum.";
RL Nature 415:497-502(2002).
DR EMBL; AL646071; CAD16422.1; -
DR InterPro; IPR003335; Sec2_SecF.
DR Pfam; PF02355; Sec2_SecF.1.
DR TIGRFAMs; TIGR00916; 2A0504s01; 1.
DR TIGRFAMs; TIGR01129; sec2; 1.
KW Complete proteome.
SQ SEQUENCE 626 AA; 66810 MW; DCC2F59A5FAC2F7B CRC64;

Alignment Scores:
Pred. No.: 4,15e-23 Length: 626
Score: 373.00 Matches: 89
Percent Similarity: 56.12% Conservative: 44
Best Local Similarity: 35.44% Mismatches: 100
Query Match: 11.26% Indels: 4
DB: 16 Gaps: 1

US-09-868-987-1 (1-1864) x Q8XVW2 (1-626)
QY 62 AACCGTGAATGGCGTATGGCTGATGAT-----GACGTTATATGCTACG 109
Db 371 AsnIleGlyIleArgMetGlyIleValleuPheGlyIleGlyValleuThr 390
QY 110 AGCCATTATTAACGTCCCATTAAGAAATCATGCGACGTGCTCGGGAATTTACCAAC 169

```

```

Db 391 ValAlaThrIleIleGlnSerIleuIleuGlySerArgPheGlnIleThrGlyIleGlySerVal 410
QY 170 CGTGAATGTGCAAACTCGCTCAGATTAAATCTGGACCGATGCTTTGTTCCCGAG 229
Db 411 GluSerAlaSerAspLeuAlaLeuLeuIleuIleuArgIleGlySerLeuAlaAlaProMetGlu 430
QY 230 GTTCTGATGAGAGACGATCTCTTGTGATCTTGGGAAAAAACAATGATACACAGCAT 269
Db 431 IleIleIleuIleuArgIleIleGlyProSerIleuGlyAlaAspAsnIleIleuIleuIleu 450
QY 290 ATTCACAGATGCTGTGGCTTGGCAATGCTTATTTGTTTGAAGCGATATTAATTAAT 349
Db 451 AsnSerAlaIleArgIlePheAlaIleIleSerValPheMetValleuIleuIleuIleu 470
QY 350 GGAGGCTCATCGCTTGGGAGCTGTTCTTCTGATCTTTGCTTATCTGGGACGCTTA 409
Db 471 PheGlyAlaPheSerValValaIleuGlyValaAsnValPheLeuIleuIleuIleu 490
QY 410 CAGTATTTGAGTGGCGACCTACCTTGTGAGAGCTCGCTGGATTTGTTGCTATGAGG 469
Db 491 SerMetLeuGlnAlaThrIleuIleuIleuProGlyIleAlaIleAlaIleAlaIleuGly 510
QY 470 ATGCGCTGATGCAATGCTTGTGATTTGTAATGCAAAAGATCCGAGAGCAATTTATGCT 529
Db 511 MetAlaIleAspAlaIleValleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 530
QY 530 CAAGCTTTAAAAATGTGAGAAAAAGATATACCAAGCTTTTGGACCATTTTGTAT 589
Db 531 AlaSerProGlnMetAlaIleAlaIleAlaGlyPheGluIleuIleuIleuIleuIleu 550
QY 590 TCTTACCTGATCAGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 649
Db 551 SerSerValThrIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 570
QY 650 AAAGGTTTCTTGTGATTTTGAATTTTGAATTTTCTTCAATGTTTACGGCTCTTTC 709
Db 571 ArgGlyPheAlaValaValaHisGlySerLeuGlyIleuIleuIleuIleuIleuIleuIleu 590
QY 710 ATGACTAAATTTTCTTCAATGCTGTGATTAATTAAGCAACCAATACACAG 760
Db 591 PheAsnArgSerLeuValaAsnLeuIleuIleuIleuIleuIleuIleuIleuIleuIleu 607

RESULT 33
Q9HX12 PRELIMINARY; PRT; 306 AA.
ID Q9HX12
AC Q9HX12;
DT 01-MAR-2001 (TReMBLrel. 16, Created)
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Secretion protein SecF.
GN SEC2 OR PA3820.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Hickey C.K., Plam X.-O.T., Erwin A.L., Mizoguchi S.D., Warren P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltzy L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.U., Coulter S.N., Folger K.R., Kas A., Laribien I.T.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olsson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen.";
RL Nature 406:959-964(2000).
DR EMBL; AE004799; AAG07207.1; -
DR InterPro; IPR003335; SecD_SecF.
DR Pfam; PF02355; SecD_SecF.1.
DR TIGRFAMs; TIGR00916; 2A0604s01; 1.
DR TIGRFAMs; TIGR00966; 3a0501s07; 1.

```

```
KW Complete proteome.
SQ SEQUENCE 306 AA; 33041 MW; 7C078C8409DF9BA2 CRC64;

Alignment Scores:
Pred. No.: 7,67e-23 Length: 306
Score: 369.50 Matches: 91
Percent Similarity: 51.69% Conservativity: 77
Best Local Similarity: 28.00% Mismatches: 102
Query Match: 11.16% Indels: 55
DB: 16 Gaps: 10

US-09-868-987-1 (1-1864) x Q9HX12 (1-306)
QY 779 TTCGTGGGATAAGCATGATTTCTTGAGAGATGCAAAAAAATTTGGGCTGTTCTTCTGGA 838
Db 10 PheMetGlyIleArgAsn-----ValAlaPhe 18
QY 839 AGTGTGTTTCTTCTAGGTTGGCTCTCGGTTTGGAGCTGGAATTCGGT-----892
Db 19 AlaValThrLeuIle---LeuThrValIleAlaLeuGlySerIrrPheThrLysGlyIle 37
QY 893 ---TTGGAATGGAATTTAAAGAGGGTATCGCTTTTACCTTTAATCCAAAGAGCATGGC 949
Db 38 AsnPheGlyLeuAspPheThrGlyThrLeuIleGluLeuThrTyrGluGlnPro---56
QY 950 ATCAGCATGTTGCTCAATCGGTGGCAAGTGTGCAATAACTACAGGAAGCTGGTCTT 1009
Db 57 ---AlaAspLeuGlyLysValArgGlyGlnLeuVal-----GlyAla 69
QY 1010 TCTTCTAGAGACTCCCGTATTCAACATTTTGGATCTTCAGAAAGATCAAAATCTATTT 1069
Db 70 GlyTyrGluAspAlaValValGlnSerPheGlyAspAlaArgAspValLeuValArgMet 89
QY 1070 AGTGATAAGCTTTAAAGCTATATAAGCAGATACGAGCTCTCTCTAAATTAACGATC 1129
Db 90 ProSerGluAspProGluLeuGlyLysValAlaThrAlaLeuGlnGlnAlaAsp---108
QY 1130 ATGAGCTGGCGTTATGTGGGATTGTGTGTGAGAAACGAGCGCTAGATTTCTCTCGGAAC 1189
Db 109 -----AlaGly 110
QY 1190 TCTAAACGAAACGAAATTTTGGTCAAGGTAAAGCAGCAACTATCGAAGAAATCGGT 1249
Db 111 AsnProAlaAsnLeuLysArgValGluTyrValGlyProGlnValGlyGluLeuArg 130
QY 1250 TATCAGCGCATCGGCTTTTAGGAGCTTTGGCAATCATCTGCTATGTGAGTTG 1309
Db 131 AspGlnGlyLeuGlyMetLeuLeuAlaLeuGlyGlyIleLeuLeuTyrValGlyPhe 150
QY 1310 CGTTTGAATGGCAATATGCTTTTCAAGTCCGTATGCGCTTTAATTCATGACCTTTGGCT 1369
Db 151 ArgPheGlnTrpLysPheAlaLeuGlyAlaIleLeuSerLeuValHisAspAlaIleIle 170
QY 1370 ACCTGTGCGCTGTTTATAGCATTCTTTTGAAGAAATCAATAGATTGCAA 1429
Db 171 ValMetGlyValLeu-----SerPhePhe-----GlnValThrPheAspLeuThr 185
QY 1430 GCCATTGTGCTTTAATGACTGTATTTGGGTATTTCATTAAACAATCTTTTCATCATTTT 1489
Db 186 ValLeuAlaValLeuAlaValValGlyTyrSerLeuAsnAspThrIleValIlePhe 205
QY 1490 GATCGTATTCGTGAAGAT-----CGCCAAGCGAACCTGTTTACCCCTATGCT 1537
Db 206 AspArgValArgGluAsnPheArgValLeuArgLysAlaAspLeuValGluAsnLeuAsn 225
QY 1538 GTTTTAGTTAATGATGCCCTTCAAAAGAGTTTCAGCCGACGGTAAGACAAAGCTACA 1597
Db 226 Ile-----SerThrSerGlnThrLeuLeuArgThrIleAlaThrSerValSer 241
QY 1598 ACTCTATCAGTTTGTGTTAATGCTTTTATAGCGGCTCTCTGCTCTTTAAATTTGCA 1657
Db 242 ThrLeuLeuAlaIleAlaAlaLeuLeuPhePheGlyGlyAspAsnLeuPheGlyPheSer 261

QY 1658 TTATATATGACCATAGGATTTCTTAGGAACCTTATCGTCTCTTATATATGACCACT 1717
Db 262 IleAlaLeuPheValGlyValMetAlaGlyThrTyrSerSerIleTyrIleAlaAsnVal 281
QY 1718 CTGTTGTTGTTTATG 1732
Db 282 ValLeuIleTrpLeu 286

RESULT 34
Q9KLO0 PRELIMINARY; PRT; 612 AA.
AC Q9KLO0;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Protein-export membrane protein SecD.
GN VCA0693.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_TaxID=666;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EL TOR N16961 / SEROTYPE O1;
RX MEDLINE=20406833; PubMed=10952301;
RA Heidelberg J.F., Eissen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA Emolaeval M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA McDonald L., Uterback T., Fleischmann R.D., Nierman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT cholerae."
RL Nature 406:477-483 (2000).
DR EMBL; AS004398; AAF96592.1; -.
DR TIGR; VCA0693; -.
DR InterPro; IPR003335; SecD_SecF.
DR Pfam; PF02355; SecD_SecF; 1.
DR TIGRFAMs; TIGR00916; 2A0604801; 1.
DR TIGRFAMs; TIGR01129; secD; 1.
KW Complete proteome.
SQ SEQUENCE 612 AA; 66894 MW; B0466FF76C175E8E CRC64;

Alignment Scores:
Pred. No.: 1.1e-22 Length: 612
Score: 368.00 Matches: 79
Percent Similarity: 59.17% Conservativity: 63
Best Local Similarity: 32.92% Mismatches: 94
Query Match: 11.11% Indels: 4
DB: 16 Gaps: 3

US-09-868-987-1 (1-1864) x Q9KLO0 (1-612)
QY 38 ACTGCTAATGACAAATATCTCGAACCGTGGATGGCTGATGCTGTAGTATGACGGT 97
Db 370 ThrValTyrArgGluTyrLysThrAsnAla-----ArgGlyGluThrValArgSerGlu 387
QY 98 TATATGCTCAGCAGCCCTATTTTAAACGTCCTCATGAAATCATGCCAGTGTCTCAGGG 157
Db 388 LysValIleSerValIleThrIleGlnSerGlnLeuGlySerGlnPheArgIleThrGly 407
QY 158 AAATTTACCCACCGTGAAGTGAGCAACTCGCTTCAGATTTTAAATCTGGAGCGATGCT 217
Db 408 AlaGlySerMetGluGluAlaGlnGlnLeuAlaLeuLeuLeuArgAlaGlySerLeuThr 427
QY 218 TTGTTCCCGAGGTTCTCAGTGAAGAGACGATCTTCTGATCTTGGGAAAAACAATGT 277
Db 428 AlaProValThrIleValGluGluArgThrIleGlyAlaSerLeuGlyGluGluAsnIle 447
QY 278 ACACAAGGCATTATCTCAGCATGCTGTGGCTTGGCAATGCTTATTGTTTTCATCAGCGTA 337
Db 448 AlaAsnGlyPheAlaAlaLeuAlaLeuGlyMetAlaMetThrLeuThrPheMetAlaLeu 467
```

[illegible]

Db	182	GlutindilyValaspertyrAlaGluGluValaMetLysAlaSerProlyserTyr-----	199
Qy	80	GCTGTAAGTATGACGGTTATATGATCGACAGCCCATTTTA-----AAGCTCCCAATG	133
Db	200	-----IleSerAlaAlaSerValaenGlnProLeuPheThrArgSpaValMetIle	216
Qy	134	AAAAATCATCCAGTGTCTTCAGAGGAATTTTACCACCGTGAAGTAGACCAATCCCTCTCA	193
Db	217	GluuSnArg-----SerGlyAspPheThrValGluGlnThrArgPheIleAlaSp	233
Qy	194	GATTTAAATCTGAGAGCATGTCCTTTTGTTCGCCAGGTTCTCAGTGAAGACGATCTCT	253
Db	234	IleLeuSnAlaGlySerLeuProAlaLysLeuAspValLeuSerSerAsnSerValSer	253
Qy	254	TCGTATCTTGAGAAAAAACAATATACACAAGCATTTATCTCAGCATGCTGTGGCTTGCA	313
Db	254	AlaSerLeuIlyGluGlnAlaMetGluArgThrValTyrAlaGlyPheIleGlyValAla	273
Qy	314	ATGCTATTTGTTTGGATGAGAGCGATATTTATTTAGATTTGGAGGGCGTATCGCTGGGAGCT	373
Db	274	LeuIlePheLeuIlyrMetLeuLeuIlyrTrArgPheMetGlyMetIleAlaThrIleThr	293
Qy	374	GTTCTCTTGATCTTTTGTTCCTTATCTGGGACGCTTACAGATTTTGGATGCGGCCTACCC	433
Db	294	LeuThrValTyrIleTyrLeuValLeuLeuIlePheAsnTyrMetAsnAlaValLeuThr	313
Qy	434	TTTGTCAGAGCATCGCTGGAGATTTGTTCTTGCTATGAGGATGGCCCGTAGATGCAAAATGTTCT	493
Db	314	LeuProGlyIleAlaIleAlaLeuIleLeuGlyValGlyMetAlaValaAspAlaSnIleIle	333
Qy	494	GTTATTCGAAAGATCCGAGAGGAATTTTATTTATGTCCAAGTCTTAAAAATCTGTAGAA	553
Db	334	ThrTyrGluArgIleLysGluGlnIleArgSerGlyLysSerIleMetSerSerPheLys	353
Qy	554	AAAGATATATACCAAGGCTTTTGGAGCCATTTTGGATTTTGAATTTGACTTACAGATTTGGCC	613
Db	354	AlaGlySerArgArgSerIleuSerThrIleLeuAspAlaAsnIleThrThrIleuAla	373
Qy	614	TCAGACACTCTTTCTTCTTCCTCAAGTACACAGGCGCTATTAAGGGTTTGCTTGACATTTGAT	673
Db	374	AlaSerValLeuTyrIleTyrGlyThrIleSerSerValGlnGlyPheAlaValMetLeuIle	393
Qy	674	TTAGAGATTTTCTCTCAATGTTTAGGGGCTTTTATCATGACAAATTTTCTTCATGCTG	733
Db	394	ValSerIleLeuMetSerPheValThrAlaValPheGlySerArgLeuLeuGlyLeu	413
Qy	734	TGATGATATAGAACCCACATACACAGTTGACATATGATGATGATGATGCGTGGGATTAAG	793
Db	414	TryAlaMetSer-----ArgThrLeuAsnLysLysTyrArgLeuPheGlyValLys	430
RESULT	36		
Q8XGX9			
ID	Q8XGX9	PRELIMINARY;	PRT; 323 AA.
AC	O8XGX9;		
DT	01-MAR-2002 (TrEMBLrel. 20, Last sequence update)		
DT	01-MAR-2002 (TrEMBLrel. 21, Last annotation update)		
DE	Protein translocase, HSP family, membrane subunit (Protein-export		
DE	membrane protein SecY).		
GN	SECF OR STM0408 OR STY0446.		
OS	Salmonella typhimurium, and		
OC	Salmonella typhi.		
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;		
OC	Salmonella.		
OX	NCBI_TaxId=602, 601;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	SPECIES=S.typhimurium; STRAIN=LT2 / SSGC1412 / ATCC 700720;		
RA	MEDLINE=21534948; PubMed=11677609;		
RA	McClintland M., Sanderson K.E., Speleth J., Clifton S.W., Latreille P.,		
RA	Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,		
RA	Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,		
RA	Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,		

RA Waterston R., Wilson R.K.;
 RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
 L72.";
 RL Nature 413:852-856(2001).
 RN [2]
 RC SEQUENCE FROM N.A.
 RC SPECIES=S.typhi; STRAIN=CT18;
 RX MEDLINE=2153494; PubMed=11677608;
 RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
 RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebaihia M.,
 RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
 RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
 RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holtroyd S., Jagels K.,
 RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
 RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
 RA Whitehead S., Barrall B.G.;
 RT "Complete genome sequence of a multiple drug resistant Salmonella
 RT enterica serovar Typhi CT18.";
 RL Nature 413:848-852(2001).
 DR EMBL; AE008714; AAU19362.1; -;
 DR EMBL; AL627266; CAD08864.1; -;
 DR InterPro; IPR003335; SecD_SecF.
 DR Pfam; PF02355; SecD_SecF; 1.
 DR TIGRFAMs; TIGR00916; 2A0604S01; 1.
 DR TIGRFAMs; TIGR00966; 3A0501S07; 1.
 DR Complete proteome.
 SQ SEQUENCE 323 AA; 35376 MW; 10D06394DF60CB86 CRC64;

Alignment Scores:
 Pred. No.: 1.85e-22 Length: 323
 Score: 365.00 Matches: 103
 Percent Similarity: 49.68% Conservative: 53
 Best Local Similarity: 32.80% Mismatches: 122
 Query Match: 11.02% Indels: 36
 DB: 16 Gaps: 11

US-09-868-987-1 (1-1864) x Q8XGX9 (1-323)

QY 794 CATGATTCTTTGAGAGATGCAAAAACCTTTGGGCT-----GTTCTCGAAGTGTTTT 847
 Db 17 TyrAepPheMetArgTrp-----AspPheTrpAlaPheGlyIleSerGlyLeuLeuLeu 34
 QY 848 CTTTATAGTTCGCTGCTCGGTTTGGAGCGCTGGAATTCGGTTTGGGAATGATTTT 907
 Db 35 IleAlaAlaIleValIleMetGlyValArgGlyPheAsn-----TrpGlyLeuAspPhe 52
 QY 908 AAAGAGGGTATGCTTTACCTTTTAATCAAAAGAGCATGCATCAGCGATGCTCTCAA 967
 Db 53 ThrGlyGlyThrValIleGluIleThrLeuGluIysProAlaGluMetAspVal----- 70
 QY 968 ATGGGTGCCAAAGTTGTGCATAAACTACAGGAAGCTGCTTCTTCTCAGAGACTTCGCT 1027
 Db 71 MetArg-----GluAlaLeuGlnIysAlaGlyTyr-----GluGluProGln 84
 QY 1028 ATTCAACATTTGATCTTCAGAAAGATCAAAATCTATTATTAGTATGAAGCTTTAAGC 1087
 Db 85 LeuGlnAsnPheGlySerSerHisAspIleMetValArgMetProProThrGluGlyGlu 104
 QY 1088 TATACTAGCAGATACAGCTCTCTCTCTAAATTAACATCATGAGCTGGCGTTATGT 1147
 Db 105 ThrGlyGlyGlnValLeuGlySerIysVal----- 114
 QY 1148 GGGATTGTGTGCAAAACAGCGCTAGATTTCTCTACGGAAACTCTAAACGAAACGCAAAA 1207
 Db 115 ---ValThrIleIleAsnGluAlaThr-----AsnGlnAsnAlaAlaValIys 129
 QY 1208 TTTTGGTCAAAGGTAAAGCAGCAAACTATCAGAAATCGTATTACGGCACCATCGGG 1267
 Db 130 ArgIleGluPheValGlyProSerValGlyAlaAspLeuAlaGlnThrGlyAlaMetAla 149
 QY 1268 CTTTATAGAGCTTTGGCAATCATCTCTCTCTATGAGCTTTGGCTTTGAATGCAATAT 1327
 Db 150 LeuLeuValAlaIleSerIleLeuValTyrValGlyPheArgPheGluTrpArgLeu 169

QY 1328 GCTTTCAGTCCGCTATGCGCTTTTAATTCATGACCTTTTGGCTACCTGCGAGTCTTGT 1387
 Db 170 AlaAlaGlyValValIleAlaLeuAlaHisAspValIleThrLeuGlyIleLeuSer 189
 QY 1388 ATAGCACATTTCTTTTGAAGAAATTCAAATAGATTTGCAAGCCATTGGTCTTTAATG 1447
 Db 190 LeuPheHis-----IleGluIleAspLeuThrIleValAlaSerLeuMet 204
 QY 1448 ACTTATTTGGGTATTTCATTAACAATCTTTGATCATCTTTTGTATTCGTTGAAGAT 1507
 Db 205 SerValIleGlyTyrSerLeuAsnAspSerIleValValSerAspArgIleArgGluAsn 224
 QY 1508 ---CGCCAAGCAACCTGTTTACCCCTATGATGATGATGATGATGATGATGATGATG 1564
 Db 225 PheArgIysIleArgArgGlyThrProTyrGluIlePhe---AsnValSerLeuThrGln 243
 QY 1565 ACCTTCAGCCGCGGTAATGACAGCACTACACTATCATCATCTATCATCTTTGTTAATGCT 1624
 Db 244 ThrLeuHisArgThrLeuIleThrSerGlyThrLeuValValIleLeuMetLeuTyr 263
 QY 1625 TTTATAGCGGCTCTCTGCTCTTTAAATTTTGCATTTATTATGACCATAGGATTTCTTA 1684
 Db 264 LeuPheGlyGlyProValLeuGluGlyPheSerLeuThrMetLeuIleGlyValSerIle 283
 QY 1685 GGAATTTATGCTCTCTTTATATGACCACTCTCTGTTGTTG 1726
 Db 284 GlyThrAlaSerSerIleTyrValAlaSerAlaLeuAlaLeu 297
 RESULT 37
 Q9K0J0 PRELIMINARY; PRT; 618 AA.
 AC Q9K0J0;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Protein-export membrane protein SecD.
 GN NMB0607.
 OS Neisseria meningitidis (serogroup B).
 OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
 OX NCBI_TaxID=491;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX STRAIN=MC58 / SEROGROUP B;
 RX MEDLINE=20175755; PubMed=10710307;
 RA Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
 RA Eisen J.A., Ketchum K.A., Hood D.W., Peden J.P., Dodson R.J.,
 RA Nelson W.C., Gwinn M.L., DeBoy R., Peterson J.D., Hickey E.K.,
 RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
 RA Mason T., Ciecko A., Parksey D.S., Blair E., Citterone H., Clark E.B.,
 RA Cotton M.D., Uterback T.R., Khouri H., Qin H., Vamathevan J.,
 RA Gill J., Scarlato V., Maignani V., Pizza M., Grandi G., Sun L.,
 RA Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;
 RT "Complete genome sequence of Neisseria meningitidis serogroup B strain
 RT MC58.";
 RL Science 287:1809-1815(2000).
 DR EMBL; AE002416; AAP41034.1; -;
 DR TIGR; NMB0607; -;
 DR InterPro; IPR003335; SecD_SecF.
 DR Pfam; PF02355; SecD_SecF; 1.
 DR TIGRFAMs; TIGR00916; 2A0604S01; 1.
 DR TIGRFAMs; TIGR01129; secD; 1.
 KW Complete proteome.
 SQ SEQUENCE 618 AA; 66847 MW; F4F805FB28B0AF10 CRC64;

Alignment Scores:
 Pred. No.: 2.4e-22 Length: 618
 Score: 364.00 Matches: 80
 Percent Similarity: 55.95% Conservative: 61
 Best Local Similarity: 31.75% Mismatches: 105
 Query Match: 10.99% Indels: 6
 DB: 16 Gaps: 3


```

Db 588 ValAsnLeuTrpTyrGlyArgArgLysLeuGln 599
RESULT 39
O9KTY6
ID O9KTY6 PRELIMINARY; PRT; 315 AA.
AC O9KTY6;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Protein-export membrane protein SecF.
GN VC0744.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_TaxID=666;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EL TOR N16961 / SEROTYPE O1;
RX MEDLINE=20406833; PubMed=10952301;
RA Heideberg J.F., Eise J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA Dodson R.J., Haft D.H., Hick E.K., Peterson J.D., Umayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA McDonald L., Uterback T., Fleischmann R.D., Nierman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT cholerae.";
RL Nature 406:477-483(2000).
DR EMBL; AB004160; AAF93909.1; -.
DR TIGR; VC0744; -.
DR InterPro; IPR003335; SecD_SecF.
DR Pfam; PF02355; SecD_SecF; 1.
DR TIGRFAMs; TIGR00916; 2A0604s01; 1.
DR TIGRFAMs; TIGR00966; 3A0501s07; 1.
KW Complete proteome.
SQ SEQUENCE 315 AA; 34329 MW; 634AF1FAFFC070BD CRC64;

Alignment Scores:
Pred. No.: 2.48e-22 Length: 315
Score: 363.50 Matches: 104
Percent Similarity: 50.62% Conservative: 59
Best Local Similarity: 32.30% Mismatches: 104
Query Match: 10.98% Indels: 55
DB: 16 Gaps: 13

US-09-868-987-1 (1-1864) x O9KTY6 (1-315)
Qy 797 GATTTCCTGAGAGGATGCAAAACCTTTGGGCTGTTCT-----GGAAAGT 841
Db 12 AspPheMetArgTrpSerLysPheAlaPheAlaLeuSerLeuValMetIleAlaAlaSer 31
Qy 842 GTTTTCTTTTAGGTGGCTGCTCGGTTTGAGGCTGGAATTCGTTTGGGAATG 901
Db 32 IlePheThrLeuSerThrLysTrpLeuAsn-----Trp-----GlyLeu 44
Qy 902 GATTTTAAAGAGGGTATGCCTTTACCTTTTAATCCAAAGAGCATGGCATGCTT 961
Db 45 AspPheThrGlyGlyThrLeuIle-----GluValGlyPheGluGlnPro 59
Qy 962 GCTCAATATGCGTGGCAAGTGTGTGATCAAACTACAGGAAGCTGGTCTTCTTCTAGAC 1021
Db 60 AlaAsnLeuGlu---GlnIleArgSerAlaLeuGluAlaLysGlyPheGly-----Asp 76
Qy 1022 TTCCTATTCAACATTTGGATCTTCAGAAAGATCAAAATCTAT----- 1066
Db 77 AlaThrValGlnAsnPheGlySerAlaArgAspValMetValArgLeuArgProArgAsp 96
Qy 1067 ---TTTAGTATAAGCTTTAAGCTTACTAAGCAGATACGACCTCTCTCTCTAAATTA 1123
Db 97 AspMetAlaGlyGluAlaLeuGly-----AsnGlnIleLeuAlaAlaIleLysGly 114
Qy 1124 ACGATCATGAGCTGGCGTATTGTGGGATTTGTGTGAGAAACAGCGCTAGATTCTCTAC 1183

```

```

Db 115 Thr-----GlyGlyAsnValGluMetArgGlyLeuGluPhe----- 126
Qy 1184 GGAACCTCTAAACGAAACGAAATTTTGTCTCAAGAGTACGACAACTATCGAAGAA 1243
Db 127 -----ValGlyProAsnValGlyAspGlu 134
Qy 1244 ATGCGTTATCAGCGACCATCGGGCTTTTAGGAGCTTTTGGCAATCATCTTGTCTATGTG 1303
Db 135 LeuThrGluAlaGlyGlyLeuAlaIleLeuValSerLeuLeuGlyLeuLeuLeuVal 154
Qy 1304 AGTTTGGCTTTGAATGGCAATATGCTTTTTCAGTCCGCTATGCGCTTTAATTTCATGACCTT 1363
Db 155 SerValArgPheGluTrpArgLeuAlaAlaGlyAlaValLeuAlaLeuAlaHisAspVal 174
Qy 1364 TTGGCTACCTGTGCAGCTGTTTATATAGCACATTTCTTTTGAAGAAAATTCAAATAGAT 1423
Db 175 IleIleThrLeuGlyIle-----PheSerIleLeuGlnIleGluValAsp 189
Qy 1424 TTGCAAGCCATTGGTCTTTAATGACTGTATTGGGGTATTATTAAACAATATCTTTGATC 1483
Db 190 LeuThrIleValAlaAlaLeuLeuThrValValGlyTyrSerLeuAsnAspThrIleVal 209
Qy 1484 ATTTTTCATCGTATTCTGTGAAGAT---CGCCAACGCAACCTGTTTACCCCTATCATGTT 1540
Db 210 ValPheAspArgIleArgGluAsnPheArgLysMetArgLysGluGluProAlaGluIle 229
Qy 1541 TTAGTTAATGATGCCCTTCAAAAGACGTTTCACGCGACGCTGTAATGACAAACAGCTTACAACT 1600
Db 230 Met---AsnSerSerIleThrGlnThrLeuSerArgThrLeuIleThrSerGlyThrThr 248
Qy 1601 CTATCAGCTTTTGTAAATGCTTTTGTATATAGGGGCTCTCTCTTTTAATTTTTCATTT 1660
Db 249 LeuPheValValIleAlaLeuPheThrGlnGlyGlyAlaMetIleHisGlyPheAlaLeu 268
Qy 1661 ATATGACCATCAGGATCTTCTAGGAACCTTATCGTCTCTTATATATGACCACTCTG 1720
Db 269 AlaLeuLeuLeuGlyIleThrValGlyThrTyrSerSerIleThrValAlaSerAlaLeu 288
Qy 1721 TTGTTG 1726
Db 289 AlaLeu 290
RESULT 40
O9PGS3 PRELIMINARY; PRT; 614 AA.
AC O9PGS3;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Protein-export membrane protein.
GN XF0225.
OS Xylella fastidiosa.
OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
OX Xylella.
OX NCBI_TaxID=2371;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=9A5C;
RX MEDLINE=20365717; PubMed=10910347;
RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
RA Alvaranga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrer H.,
RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,
RA Facinani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,

```



```

Qy 461 GCTATGGGATGGCCGTAGATCAAAATGTTCTTGTATTTCGAAAGAAATCCGAGAGGAATTT 520
Db 513 ThrLeuAlaValAlaAspAlaAsnValLeuIleAsnGluArgIleLysGluGluLeu 532
Qy 521 TTATTGTCTCAAGCTTAAAGAAATCTCTAGAAAAGGATATACCAAGCGCTTTTCGAGCC 580
Db 533 SerAsnGlyArgSerValGlnGlnAlaIleAsnGluGlyTyrAlaGlyAlaPheSer 552
Qy 581 ATTTTGTATTCTAATCTAGTACATGATTTGGCCCTCAGCACTCTTTTCTCTCTAGATACA 640
Db 553 IlePheAspAlaAsnIleThrThrLeuIleLysValIleIleLeuTyrAlaValGlyThr 572
Qy 641 GGGCCTATTAAGGTTTGGCTTTCATGATTTAGGAATTTTCTCTCAATGTTTACG 700
Db 573 GlyAlaIleLysGlyPheAlaIleThrThrGlyIleGlyValAlaIleThrSerMetPheThr 592
Qy 701 GCTCTTTTCATGACTAAA 718
Db 593 AlaIleValGlyThrArg 598

RESULT 42
Q8ZRD7 PRELIMINARY; PRT; 615 AA.
AC Q8ZRD7;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DE Preprotein translocase, IISF family, part of the channel.
GN SEC OR STM0407
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=L72 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Flores L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
L72."
RL Nature 413:852-856(2001).
DR EMBL; AE008714; AAL19361.1; -.
DR InterPro; IPR003335; SecD_SecF.
DR InterPro; IPR000508; SigPase.
DR Pfam; PF02335; SecD_SecF; 1.
DR TIGRFAMs; TIGR00916; 2A0604s01; 1.
DR TIGRFAMs; TIGR01129; secD; 1.
DR PROSITE; PS00761; SPASE_1_3; UNKNOWN_1.
KW Complete proteome.
SQ SEQUENCE 615 AA; 66587 MW; 38383148F3916B33 CRC64;

Alignment Scores:
Pred. No.: 1.04e-21 Length: 615
Score: 356.50 Matches: 83
Percent Similarity: 59.75% Conservative: 61
Best Local Similarity: 34.44% Mismatches: 94
Query Match: 10.76% Indels: 3
DB: 16 Gaps: 2

US-09-868-987-1 (1-1864) x Q8ZRD7 (1-615)
Qy 44 AATGACAAATATCTGAAACCGTGGATGGCGTATGCGTATGATTCGCGTATATG 103
Db 375 SerGlyLysLysAspAlaAsn-----GlyArgAlaValLeuValLysGlnGluVal 392
Qy 104 GTCAGAGCCCTATTTTAAACGTCCTCCATGGAATATCATGCGATGCTCAGGGAATTT 163
Db 393 IleAsnIleAlaAsnIleGlnSerArgLeuGlyAsnSerPheArgIleThrGlyIleSer 412

```

```

Qy 164 ACCACCGTGAAGTGAAGCAAACTCGCCTCAGATTTAAATCTGGAGCGATGCTTTTGT 223
Db 413 AsnProAsnGluAlaArgGlnLeuSerLeuLeuLeuArgAlaGlyAlaLeuIleAlaPro 432
Qy 224 CCCAGGTTCTCAGTGAAGAGACCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 283
Db 433 IleGlnIleValGluGluArgThrIleGlyProThrLeuGlyMetGlnAsnIleLysGln 452
Qy 284 GGCATTATCTCAGCATGCTGGGCTTGCAATGCTTATTGTTTGTATGAGCGTATATTAT 343
Db 453 GlyLeuGluAlaCysLeuAlaGlyLeuValSerIleLeuPheMetIlePhePheThr 472
Qy 344 AGATTGGAGCGCTCATCGCTTCGGGAGCTGTTCTCTCTGAATCTTTTCTCTATCTGGCA 403
Db 473 LysLysPheGlyLeuIleAlaThrSerAlaLeuValAlaAsnLeuValIleValGly 492
Qy 404 GCTCTACAGTATTG---GATGCGCCACTCACCTTGTCTGAGCACTCGCTGGGATGTTCT 460
Db 493 IleMetSerLeuLeuProGlyAlaThrLeuSerMetProGlyIleAlaGlyIleValLeu 512
Qy 461 GCTATGGGATGGCCGTAGATGCAAAATGTTCTTCTTATTCGAAAGAAATCCGAGAGGAATTT 520
Db 513 ThrLeuAlaValAlaAspAlaAsnValLeuIleAsnGluArgIleLysGluGluLeu 532
Qy 521 TTATTGTCTCAAGCTTAAAGAAATCTGTAGAAAAGGATATACCAAGGCTTTTGGAGCC 580
Db 533 SerAsnGlyArgThrValGlnGlnAlaIleAsnGluGlyTyrAlaGlyAlaPheSerSer 552
Qy 581 ATTTTGTATCTAATCTGACTACATGATTTGGCCCTCAGCACTCTTTTCTCTCTAGATACA 640
Db 553 IlePheAspAlaAsnIleThrThrLeuIleLysValIleIleLeuTyrAlaValGlyThr 572
Qy 641 GGGCCTATTAAGGTTTGGCTTTCATGATTTAGGAATTTTCTCTCTCAATGTTTACG 700
Db 573 GlyAlaIleLysGlyPheAlaIleThrThrGlyIleGlyValAlaIleThrSerMetPheThr 592
Qy 701 GCTCTTTTCATGACTAAATTTTCTCATGCTGTGGATGAATAAGACCAACATACACACAG 760
Db 593 AlaIleIleGlyThrArgAlaIleValAsnLeuLeuTyrGlyGlyLysArgValThrLys 612
Qy 761 TTG 763
Db 613 Leu 613

RESULT 43
Q8Z8X9 PRELIMINARY; PRT; 615 AA.
ID Q8Z8X9
AC Q8Z8X9;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Protein-export membrane protein secD.
GN STY0445.
OS Salmonella typhi.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OX NCBI_TaxID=601;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CT18;
RX MEDLINE=21534947; PubMed=11677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,
RA Baker S., Basham D., Brooks R.M., Chillingworth T., Connor P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrrell B.G.;
RT "Complete genome sequence of a multiple drug resistant Salmonella
enterica serovar Typhi CT18."
RL Nature 413:848-852(2001).

```

DR EMBL: AL627266; CND08863.1; -
 DR InterPro: IPR003335; SecD_SecF.
 DR InterPro: IPR000508; SigPase.
 DR Pfam: PF02355; SecD_SecF.1.
 DR TIGRPFAM: TIGR00916; 2A0604s01.1.
 DR TIGRPFAM: TIGR01129; secD.1.
 DR PROSITE: PS00761; SPASE_1.3; UNKNOWN_1.
 DR Complete proteome.
 SQ SEQUENCE 615 AA; 66605 MW; 81370B11FD5983F9 CRC64;

Alignment Scores:

Pred. No.:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
Score:	1,04e-21	615	356.50	61	94	2
Percent Similarity:	59.75%					
Best Local Similarity:	34.44%					
Query Match:	10.76%					

US-09-868-987-1 (1-1864) x Q828X9 (1-615)

QY 44 AATGACATATATTCGCAACCGTGATGCGTATGCTGATGATTGACGTTATATG 103
 Db 375 SerLysLysAspAlaAsn-----GlyArgAlaValLeuValLysGlnGluVal 392
 QY 104 GTCAGCAGCCATATTTAAAGTCCCATTTGAAAATCATGCGACATGCTCGGAAATTT 163
 Db 393 IleAsnIleAlaAsnIleGlnSerArgLeuGlyAsnSerPheArgIleThrGlyIleSer 412
 QY 164 ACCACCGTGAAAGTGACCAACCTGCGATGATTTAAATCTGAGAGCATGCTTTGTT 223
 Db 413 AsnProAsnGlnAlaArgGlnLeuSerLeuLeuAsnArgAlaGlyAlaLeuIleAlaPro 432
 QY 224 CCCAGGTTCTCAGTGAAGAGACGATCTCTTGATCTTGGAAGAAAACAATGTACAA 283
 Db 433 IleGlnIleValGlnGluArgThrIleGlyProThrLeuGlnMetGlnAsnIleLysGln 452
 QY 284 GGCATTATTCAGATGCTGCTGCTGCGATGCTTATGTTTGATGAGGATATATAT 343
 Db 453 GlyLeuGlnAlaCysLeuAlaGlyLeuValIleSerIleLeuPheMetIlePhePheTyr 472
 QY 344 AGATTGGAGGCGTCATGCTGCGAGAGCTGCTTCTGATCTTGATCTTTGCTTATCTG 403
 Db 473 LysLysPheGlyLeuIleAlaThrSerIleAlaValAlaAsnLeuValLeuIleValGly 492
 QY 404 GCTCTACAGTATTTG--GATGCGCCACTCACTTTCAGACCTGCTGGATTTGTT 460
 Db 493 IleMetSerLeuLeuProGlyAlaThrLeuSerMetProGlyIleAlaGlyIleValLeu 512
 QY 461 GCTATGGGATGCGCGTCATGCTGCGAGAGCTGCTTCTGATCTTGATCTTTGCTTATCTG 520
 Db 513 ThrLeuAlaValAlaValAspAlaAsnValLeuIleAsnGlnArgIleLysGlnGluLeu 532
 QY 521 TTATGTCGCAAGCTTAAATAATCTGTAGAAAAGATATACCAAGCTTTGAGACC 580
 Db 533 SerAsnGlyArgThrValGlnGlnAlaIleAsnGlnGlyTyrAlaGlyAlaPheSerSer 552
 QY 581 ATTTTGAATCTAATCTGATCAAGTATGCGCTCAGACCTTTCTTCTCTCAATACA 640
 Db 553 IlePheAspAlaAsnIleThrThrLeuIleLysValIleIleLeuValAlaGlyThr 572
 QY 641 GGGCTTAAAGGTTGCTTGAATGATTTTGAATTTTCTCTCAATGTTATG 700
 Db 573 GlyAlaIleLysGlyPheAlaIleThrThrArgIleGlyValAlaThrSerMetPheThr 592
 QY 701 GCTCTTTCATGATTAATTTTCTTCTGATGATGATGATTAAGACCAATCATACACAG 760
 Db 593 AlaIleIleGlyThrArgAlaIleValAsnLeuLeuTyrGlyGlyLysArgValThrLys 612
 QY 761 TTG 763
 Db 613 Leu 613
 RESULT 44

Q9XBN2
 ID Q9XBN2 PRELIMINARY; PRI; 615 AA.

AC Q9XBN2; 01-NOV-1999 (TREMBLrel. 12, Created)
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE SecD protein.

GN SEC D
 OS Enterobacter aerogenes (Enterobacter aerogenes).
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Enterobacter.
 OX NCBI_TaxID=548;
 RN [1]
 RP SEQUENCE FROM N.A.

RA Kuhn F.C., Gies A.J., Smeltzer M., Crupper S.S., Sobieski R.J.;
 RT "Identification of the secD gene of Enterobacter aerogenes."
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.

DR EMBL: AF163861; AAD4348.1; -
 DR TIGRPFAM: TIGR00916; 2A0604s01.1.
 DR InterPro: IPR003335; SecD_SecF.
 DR Pfam: PF02355; SecD_SecF.1.
 DR TIGRPFAM: TIGR00916; 2A0604s01.1.
 DR PROSITE: PS00761; SPASE_1.3; UNKNOWN_1.
 SQ SEQUENCE 615 AA; 66744 MW; 6BA2C06542F715D3 CRC64;

Alignment Scores:

Pred. No.:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
Score:	1.14e-21	615	356.00	77	55	6
Percent Similarity:	59.74%					
Best Local Similarity:	37.20%					
Query Match:	10.75%					

US-09-868-987-1 (1-1864) x Q9XBN2 (1-615)

QY 116 ATTTTAAAGCTCCCA-----TTGAAAATCATGCGACGATGCTGAGGAAA 160
 Db 392 ValIleAsnIleAlaAsnIleGlnSerArgLeuGlyAsnSerPheArgIleThrGlyIle 411
 QY 161 TTAAACCCAGCGTGAAGTGACCAACCTGCGATGATTTAAATCTGAGAGCATGCTTTT 220
 Db 412 SerAsnProThrGlnAlaArgGlnLeuSerLeuLeuAsnArgAlaGlyAlaLeuIleAla 431
 QY 221 GTTCCGAGGTTCTCAGTGAAGAGACGATCTCTTGATCTTTGGAAGAAAACAATGTAC 280
 Db 432 ProIleGlnIleValGlnGluArgThrIleGlyProThrLeuGlnMetGlnAsnIleLys 451
 QY 281 CAAGGCTTATCTCAGATCTGCTGCGAGAGCTGCTTCTGATCTTTGATGAGCGTATAT 340
 Db 452 GlnIleLeuGlnAlaCysLeuAlaGlyLeuValIleSerIleLeuPheMetIlePhePhe 471
 QY 341 TATGATTTGAGAGCGTCATGCTGCGAGAGCTGCTTCTGATCTTTGATCTTTGCTTATCTG 400
 Db 472 TyrLysLysPheGlyLeuIleAlaThrSerAlaLeuIleAlaAsnLeuValLeuIleVal 491
 QY 401 GCAGCTTACAGTATTTG--GATGCGCCACTCACTTTCAGACCTGCTGGATTTGTT 457
 Db 492 GlyIleMetSerLeuLeuProGlyAlaThrLeuThrMetProGlyIleAlaGlyIleVal 511
 QY 458 CTTCGCTTGGAGATGCGCGTCATGCTGCGAGAGCTGCTTCTGATCTTTGATGAGCGTATAT 517
 Db 512 LeuThrLeuAlaValAlaValAspAlaAsnValLeuIleAsnGlnArgIleLysGlnGlu 531
 QY 518 TTTTATGTCGCAAGCTTAAATAATCTGTAGAAAAGATATACCAAGCTTTTGA 577
 Db 532 LeuSerAsnGlyArgThrValGlnGlnAlaIleAspGlnGlyTyrLysGlyAlaPheSer 551
 QY 578 GCAATTTTGAATCTAATCTGATCAAGTATGCGCTCAGACCTTTCTTCTCTCAATGAT 637
 Db 552 SerIlePheAspAlaAsnValThrThrLeuIleLysValIleIleLeuTyrAlaValGly 571
 QY 638 ACAGGCGCTATTAAGGTTTGGCTTGAATGATTTTGAATTTTCTCTCAATGTTT 697

```

|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
Db 572 ThrGlyAlaIleuYsGlyPheAlaIleThrThrGlyIleGlyIleAlaThrSerMetPhe 591
Qy 698 ACGGCTCTTTTCATGACTAAA 718
Db 592 ThrAlaIleValGlyThrArg 598
RESULT 45
Q9WZM4 PRELIMINARY; PRT; 465 AA.
AC Q9WZM4;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DE 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Protein-export membrane protein SEC2, putative.
GN TW0860.
OS Thermotoga maritima.
OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
OX NCBI_TaxID=2336;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=MSB8 / DSM 3109;
RX MEDLINE=99287316; PubMed=10360571;
RA Nelson K.E., Clayton R.A., Gill S.R., Swinn M.L., Dodson R.J.,
Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RA "Evidence for lateral gene transfer between Archaea and Bacteria from
RT genome sequence of Thermotoga maritima.";
RL Nature 399:323-329(1999).
DR EMBL; AE001752; AAD35942.1; -
DR TIGR; TM0860; -
DR InterPro; IPR001036; Acrflavin res.
DR InterPro; IPR003335; Sec2_SecF.
DR Pfam; PF02355; Sec2_SecF; 1.
DR PRINTS; PRO0702; ACRIFLAVINRP.
DR TIGRFAMs; TIGR00916; 2A0604e01; 1.
DR TIGRFAMs; TIGR01129; sec2; 1.
KW Complete proteome.
SQ SEQUENCE 465 AA; 50931 MW; 6CE9545B3EFD97F9 CRC64;

```

Alignment Scores:

```

Pred. No.: 1.23e-21 Length: 465
Score: 355.50 Matches: 78
Percent Similarity: 56.84% Conservative: 55
Best Local Similarity: 33.33% Mismatches: 80
Query Match: 10.73% Indels: 21
DB: 16 Gaps: 5

```

US-09-868-987-1 (1-1864) x Q9WZM4 (1-465)

```

Qy 47 GGACAATATTCGCAACCGTGGATGCGGTATGCTGTAGTATGATGACGGTTATATG--- 103
Db 219 GlyThrTyAspProLysLys-----ArgLeuAlaIleValLeuAspValValGln 236
Qy 104 GTCAGCAGCCCTATTTTAAACGTCCTCCATGAAATCATGCCAGTGTCTCAGGAAATTT 163
Db 237 PheValGlyGlnValValAlaIleIleThrAspGlyLysAlaGluIleThrGlyAsnPhe 256
Qy 164 ACCCAGCGTGAAGTACGCAAACTCGCTCAGATTTAAATCTGGAGCGATG----- 214
Db 257 SerLeuGluGluAlaLysGlnLeuAlaIleLeuArgSerGlyAlaLeuProAlaArg 276
Qy 215 -----TCTTTTGTCCCGAGGTCTCAGTGAAGACGATCTCTTCT 256
Db 277 LeuValLysThrSerSerGlyTyrValAlaProLeuLeuGlyArgAspValIleAspAla 296
Qy 257 GATCTTGGAAAAACAATGTACACAGGCATTTATCTCAGCATGCTGTGGCTTGGCAATG 316
Db 297 SerLeu-----LysAlaGlyIleIle-----GlyLeuIleLeu 307

```

```

Qy 317 CTATTGTTTGTAGAGCGTATATATATAGATTGTTGGAGCGCTCATCGCTTCGGGAGCTGTT 376
Db 308 ValLeuValTyMetIleIleTyTyArgThrMetGlyIleValAlaAspLeuAlaLeu 327
Qy 377 CTCTGAATCTTTTGTCTTATCTGGCAGCTCTACAGTATTGGATGGCCCACTCACCTTG 436
Db 328 IleTyAsnThrValLeuLeuLeuGlyValMetAlaAlaGlyLysPheIleLeuThrLeu 347
Qy 437 TCAGGACTCGCTGGGATTGTTCTTGTATCGGATGGCGCTAGATGCAAAATGTTCTTGT 496
Db 348 ProGlyIleAlaGlyIleIleLeuThrIleGlyThrThrValAspGlyAsnValIleIle 367
Qy 497 TTCGAAAGAATCCGAGAGGAATTTTATCTCTCAAGTCTTAAAAAATCTGTAGAAAAA 556
Db 368 TyrGluArgIleLysGluGluMetArgLeuGlyLysProValLysThrSerIleAlaAla 387
Qy 557 GGATATACCAAGGCTTTTGGAGCGCATTTTGTATCTTAACCTGACTACAGTATTGGCCTCA 616
Db 388 GlyPheAspArgSerLeuSerThrIleLeuAspAlaAsnIleThrThrIleLeuThrGly 407
Qy 617 GCATCTCTTTTCTCTCTAGATACAGGCGCTTATTAAAGGTTTGTCTTGACATTTGATTTA 676
Db 408 LeuIleLeuTyTyPheGlyThrGlyThrIleLysGlyPheAlaIleThrLeuIleIle 427
Qy 677 GGAATTTCTCTTCAATGTTTACGGCTCTTTTCATGACTAAA 718
Db 428 GlyValLeuGlySerIlePheValAsnLeuValPheSerArg 441

```

Search completed: December 17, 2002, 12:38:43
Job time : 109.5 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

.Run on: December 17, 2002, 09:09:46 ; Search time 64 Seconds

(without alignments)

8931.961 Million cell updates/sec

Title: US-09-868-987-1

Perfect score: 1864

Sequence: 1 atggaactccgcataattgtc.....atgacaaattcagataatgc 1864

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 441362 seqs, 1533381 residues

Word size : 0

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : Issued Patents NA.*

1: /cgn2_6/ptodata/1/ina/5A.COMB.seq.*

2: /cgn2_6/ptodata/1/ina/5B.COMB.seq.*

3: /cgn2_6/ptodata/1/ina/6A.COMB.seq.*

4: /cgn2_6/ptodata/1/ina/6B.COMB.seq.*

5: /cgn2_6/ptodata/1/ina/pCTUS.COMB.seq.*

6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	19	1.0	2972	2	US-08-720-484A-3
C 2	19	1.0	2972	3	US-08-953-822A-3
C 3	19	1.0	2972	4	US-09-398-239-3
C 4	18	1.0	378	4	US-09-134-001C-182
C 5	18	1.0	1315	2	US-08-343-101A-8
C 6	18	1.0	1315	3	US-09-183-688-8
C 7	18	1.0	2167	4	US-09-325-932A-102
C 8	18	1.0	2259	1	US-08-420-235B-20
C 9	18	1.0	2259	3	US-08-793-624-20
C 10	18	1.0	2259	5	PCT-US95-10194-20
C 11	18	1.0	3715	4	US-09-085-199B-44
C 12	18	1.0	4796	4	US-09-085-199B-3
C 13	18	1.0	20710	1	US-08-420-235B-1
C 14	18	1.0	20710	3	US-08-793-624-1
C 15	18	1.0	20710	5	PCT-US95-10194-1
C 16	18	1.0	28473	4	US-08-961-527-83
C 17	18	1.0	35100	2	US-08-770-379-18
C 18	18	1.0	35100	4	US-08-757-669A-18
C 19	18	1.0	35100	4	US-09-230-371A-18
C 20	18	1.0	36159	4	US-09-749-588-3
C 21	18	1.0	40138	3	US-09-090-753-12
C 22	17	0.9	800	1	US-08-756-771-2
C 23	17	0.9	800	2	US-09-096-571-2
C 24	17	0.9	800	4	US-09-309-320-2
C 25	17	0.9	926	3	US-08-617-860B-21
C 26	17	0.9	1137	4	US-09-045-583-6
C 27	17	0.9	1137	4	US-09-534-185-6

Sequence 4, Appli
Sequence 4, Appli
Sequence 79, Appli
Sequence 116, App
Sequence 165, App
Sequence 17, Appli
Sequence 1, Appli
Sequence 1, Appli
Sequence 8, Appli
Sequence 46, Appli
Sequence 46, Appli
Sequence 189, App
Sequence 167, App
Sequence 3, Appli
Sequence 3, Appli
Sequence 105, App
Sequence 3, Appli
Sequence 3, Appli
Sequence 5, Appli
Sequence 10, Appli
Sequence 3, Appli
Sequence 1, Appli
Sequence 14, Appli
Sequence 16, Appli
Sequence 99, Appli
Sequence 99, Appli
Sequence 99, Appli
Sequence 99, Appli
Sequence 191, App
Sequence 215, App
Sequence 7, Appli
Sequence 80, Appli
Sequence 80, Appli
Sequence 80, Appli
Sequence 80, Appli
Sequence 80, Appli
Sequence 80, Appli
Sequence 88, Appli
Sequence 88, Appli
Sequence 88, Appli
Sequence 170, App
Sequence 67, Appli
Sequence 71, Appli
Sequence 71, Appli
Sequence 19, Appli
Sequence 19, Appli
Sequence 67, Appli
Sequence 67, Appli
Sequence 67, Appli
Sequence 69, Appli
Sequence 69, Appli
Sequence 69, Appli
Sequence 25, Appli
Sequence 26, Appli
Sequence 26, Appli
Sequence 27, Appli
Sequence 27, Appli
Sequence 19, Appli
Sequence 1, Appli
Sequence 28, Appli
Sequence 73, Appli
Sequence 73, Appli
Sequence 73, Appli
Sequence 380, App
Sequence 381, App


```
977 14 0 8 1242 4 US-08-174-739A-1 Sequence 1, Appl1
978 14 0 8 1242 4 US-08-434-256-1 Sequence 1, Appl1
979 14 0 8 1246 1 US-08-446-777-3 Sequence 3, Appl1
980 14 0 8 1247 4 US-09-399-913-15 Sequence 15, Appl1
981 14 0 8 1247 4 US-09-298-731-15 Sequence 15, Appl1
982 14 0 8 1257 4 US-09-064-411A-6 Sequence 6, Appl1
983 14 0 8 1257 4 US-09-134-001C-916 Sequence 916, App
984 14 0 8 1272 4 US-09-134-001C-2358 Sequence 2358, Ap
985 14 0 8 1276 4 US-09-064-411A-29 Sequence 29, Appl1
986 14 0 8 1278 4 US-09-347-803-3 Sequence 306, App
987 14 0 8 1279 1 US-08-146-010A-4 Sequence 4, Appl1
988 14 0 8 1279 1 US-08-674-168-9 Sequence 9, Appl1
989 14 0 8 1287 1 US-08-064-121-3 Sequence 3, Appl1
990 14 0 8 1287 1 US-08-478-015-3 Sequence 3, Appl1
991 14 0 8 1287 3 US-08-475-975-3 Sequence 3, Appl1
992 14 0 8 1287 3 US-09-084-889-3 Sequence 3, Appl1
993 14 0 8 1287 3 US-09-084-889-3 Sequence 3, Appl1
994 14 0 8 1295 1 US-08-245-295-6 Sequence 6, Appl1
995 14 0 8 1295 1 US-08-481-130-6 Sequence 6, Appl1
996 14 0 8 1295 1 US-08-656-984A-6 Sequence 6, Appl1
997 14 0 8 1295 1 US-08-485-604-6 Sequence 6, Appl1
998 14 0 8 1295 2 US-08-487-595-6 Sequence 6, Appl1
999 14 0 8 1295 3 US-08-863-790-25 Sequence 25, Appl1
1000 14 0 8 1295 3 US-08-296-749-25 Sequence 25, Appl1
```

ALIGNMENTS

```
RESULT 1
US-08-720-484A-3/C
; Sequence 3, Application US/08720484A
; Patent No. 5990281
; GENERAL INFORMATION:
; APPLICANT: Desauvage, Frederic
; APPLICANT: Rosenthal, Arnon
; APPLICANT: Stone, Donna
; TITLE OF INVENTION: Vertebrate Smoothened Proteins
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatIn (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/720,484A
; FILING DATE: 30-Sep-1996
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Svoboda, Craig G.
; REGISTRATION NUMBER: 39,044
; REFERENCE/DOCKET NUMBER: P1050
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650/225-1469
; TELEFAX: 650/952-9881
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2972 base pairs
; TYPE: Nucleic Acid
; STRANDEDNESS: Single
; TOPOLOGY: Linear
; US-08-720-484A-3

Query Match 1.0%; Score 19; DB 2; Length 2972;
Best Local Similarity 100.0%; Pred. No. 6.2;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1280 TTGGCAATCATCTTGCTCT 1298
Db 1724 TTGGCAATCATCTTGCTCT 1706
```

```
RESULT 2
US-08-953-823A-3/C
; Sequence 3, Application US/08953823A
; Patent No. 616958
; GENERAL INFORMATION:
; APPLICANT: Desauvage, Frederic
; APPLICANT: Rosenthal, Arnon
; APPLICANT: Stone, Donna
; TITLE OF INVENTION: Vertebrate Smoothened Proteins
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatIn (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/953,823A
; FILING DATE: 30-Sep-1997
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/027070
; FILING DATE: 30-Sep-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Svoboda, Craig G.
; REGISTRATION NUMBER: 39,044
; REFERENCE/DOCKET NUMBER: P1050R1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650/225-1489
; TELEFAX: 650/952-9881
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2972 base pairs
; TYPE: Nucleic Acid
; STRANDEDNESS: Single
; TOPOLOGY: Linear
; US-08-953-823A-3

Query Match 1.0%; Score 19; DB 3; Length 2972;
Best Local Similarity 100.0%; Pred. No. 6.2;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1280 TTGGCAATCATCTTGCTCT 1298
Db 1724 TTGGCAATCATCTTGCTCT 1706

RESULT 3
US-09-398-239-3/C
; Sequence 3, Application US/09398239
; Patent No. 6407216
; GENERAL INFORMATION:
; APPLICANT: Desauvage, Frederic
; APPLICANT: Rosenthal, Arnon
; APPLICANT: Stone, Donna
; TITLE OF INVENTION: Vertebrate Smoothened Proteins
; FILE REFERENCE: P1050D1
; CURRENT APPLICATION NUMBER: US/09/398,239
; CURRENT FILING DATE: 1999-09-20
; EARLIER APPLICATION NUMBER: US 08/720,484
; EARLIER FILING DATE: 1996-09-30
```

; NUMBER OF SEQ ID NOS: 5
 ; SEQ ID NO 3
 ; LENGTH: 2972
 ; TYPE: DNA
 ; ORGANISM: Homo Sapiens
 US-09-398-239-3

Query Match 1.0%; Score 19; DB 4; Length 2972;
 Best Local Similarity 100.0%; Pred. No. 6.2;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1280 TTGGCAATCATCTGCTCT 1298
 |||||
 Db 1724 TTGGCAATCATCTGCTCT 1706

RESULT 4

US-09-134-001C-182
 ; Sequence 182, Application US/09134001C
 ; Patent No. 6380370

GENERAL INFORMATION:

; APPLICANT: Lynn Doucette-Stamm et al
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
 ; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: CTC-007
 ; CURRENT APPLICATION NUMBER: US/09/134,001C
 ; CURRENT FILING DATE: 1998-08-13
 ; PRIOR APPLICATION NUMBER: US 60/064,964
 ; PRIOR FILING DATE: 1997-11-08
 ; PRIOR APPLICATION NUMBER: US 60/055,779
 ; PRIOR FILING DATE: 1997-08-14
 ; NUMBER OF SEQ ID NOS: 5674
 ; SEQ ID NO 182
 ; LENGTH: 378
 ; TYPE: DNA
 ; ORGANISM: Staphylococcus epidermidis
 US-09-134-001C-182

Query Match 1.0%; Score 18; DB 4; Length 378;
 Best Local Similarity 100.0%; Pred. No. 18;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 266 AAAAAACAATGTACAA 283
 |||||
 Db 211 AAAAAACAATGTACAA 228

RESULT 5

US-08-343-101A-8
 ; Sequence 8, Application US/08343101A
 ; Patent No. 5830759

GENERAL INFORMATION:

; APPLICANT: Chang, Yuan
 ; APPLICANT: Moore, Patrick S.
 ; TITLE OF INVENTION: Unique Associated Kaposi's Sarcoma
 ; TITLE OF INVENTION: Virus Sequences And Uses Thereof
 ; NUMBER OF SEQUENCES: 22
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Cooper & Dunham
 ; STREET: 1185 Avenue of the Americas
 ; CITY: New York
 ; STATE: New York
 ; COUNTRY: U.S.A.
 ; ZIP: 10036
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.24
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/343,101A
 ; FILING DATE:
 ; CLASSIFICATION: 514

; ATTORNEY/AGENT INFORMATION:
 ; NAME: White Esq., John P.
 ; REGISTRATION NUMBER: 28,678
 ; REFERENCE/DOCKET NUMBER: 45185-A
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 212-278-0400
 ; TELEFAX: 212-391-0526
 ; INFORMATION FOR SEQ ID NO: 8:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1315 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: DNA (genomic)
 ; HYPOTHETICAL: N
 ; ANTI-SENSE: N
 US-08-343-101A-8

Query Match 1.0%; Score 18; DB 2; Length 1315;
 Best Local Similarity 100.0%; Pred. No. 19;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 396 TCTGGCAGCTCTACAGT 413
 |||||
 Db 526 TCTGGCAGCTCTACAGT 543

RESULT 6

US-09-183-688-8
 ; Sequence 8, Application US/09183688
 ; Patent No. 6093550

GENERAL INFORMATION:

; APPLICANT: Chang, Yuan
 ; APPLICANT: Moore, Patrick S.
 ; TITLE OF INVENTION: Unique Associated Kaposi's Sarcoma
 ; TITLE OF INVENTION: Virus Sequences And Uses Thereof
 ; NUMBER OF SEQUENCES: 22
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Cooper & Dunham
 ; STREET: 1185 Avenue of the Americas
 ; CITY: New York
 ; STATE: New York
 ; COUNTRY: U.S.A.
 ; ZIP: 10036
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.24
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/183,688
 ; FILING DATE:
 ; CLASSIFICATION:

; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/343,101
 ; FILING DATE:
 ; ATTORNEY/AGENT INFORMATION:

; NAME: White Esq., John P.
 ; REGISTRATION NUMBER: 28,678
 ; REFERENCE/DOCKET NUMBER: 45185-A
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 212-278-0400
 ; TELEFAX: 212-391-0526
 ; INFORMATION FOR SEQ ID NO: 8:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1315 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: DNA (genomic)
 ; HYPOTHETICAL: N
 ; ANTI-SENSE: N
 US-09-183-688-8

Query Match 1.0%; Score 18; DB 3; Length 1315;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 396 TCTGGGAGCTCTACAGT 413
DB 526 TCTGGGAGCTCTACAGT 543

RESULT 7

US-09-325-932A-102
; Sequence 102, Application US/09325932A
; Patent No. 6451604
; GENERAL INFORMATION:
; APPLICANT: Flinn, Barry
; APPLICANT: Lasham, Annette
; TITLE OF INVENTION: Compositions affecting programmed cell
; TITLE OF INVENTION: death and their use in the modification of forestry plant devel
; FILE REFERENCE: 1022
; CURRENT APPLICATION NUMBER: US/09/325,932A
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 206
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 102
; LENGTH: 2167
; TYPE: DNA
; ORGANISM: Pinus radiata
US-09-325-932A-102

Query Match 1.0%; Score 18; DB 4; Length 2167;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 655 GTTGCTTGACATTGAT 672
DB 1610 GTTGCTTGACATTGAT 627

RESULT 8

US-08-420-235B-20
; Sequence 20, Application US/08420235B
; Patent No. 5801042
; GENERAL INFORMATION:
; APPLICANT: Chang, Yuan
; APPLICANT: Moore, Patrick S.
; TITLE OF INVENTION: UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS
; TITLE OF INVENTION: SEQUENCES AND USES THEREOF
; NUMBER OF SEQUENCES: 47
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/420,235B
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 45185-B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; INFORMATION FOR SEQ ID NO: 20:

SEQUENCE CHARACTERISTICS:

LENGTH: 2259 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: N
ANTI-SENSE: N
FEATURE:
NAME/KEY: CDS
LOCATION: 1..2259
OTHER INFORMATION:

US-08-420-235B-20

Query Match 1.0%; Score 18; DB 1; Length 2259;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 396 TCTGGGAGCTCTACAGT 413
DB 1194 TCTGGGAGCTCTACAGT 1211

RESULT 9

US-08-793-624-20
; Sequence 20, Application US/08793624C
; Patent No. 6150093
; GENERAL INFORMATION:
; APPLICANT: Chang, Yuan
; APPLICANT: Moore, Patrick S.
; TITLE OF INVENTION: Unique Associated Kaposi's Sarcoma Virus Sequences And
; TITLE OF INVENTION: Uses Thereof
; FILE REFERENCE: 45185-C-PCT-US/JPW
; CURRENT APPLICATION NUMBER: US/08/793,624C
; CURRENT FILING DATE: 1997-02-18
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 20
; LENGTH: 2259
; TYPE: DNA
; ORGANISM: Kaposi's sarcoma-associated herpesvirus
US-08-793-624-20

Query Match 1.0%; Score 18; DB 3; Length 2259;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 396 TCTGGGAGCTCTACAGT 413
DB 1194 TCTGGGAGCTCTACAGT 1211

RESULT 10

PCT-US95-10194-20
; Sequence 20, Application PC/TUS9510194
; GENERAL INFORMATION:
; APPLICANT: The Trustees of Columbia University in the City of New York
; APPLICANT: City
; TITLE OF INVENTION: UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS
; TITLE OF INVENTION: SEQUENCES AND USES THEREOF
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.25

;/ CURRENT APPLICATION DATA:
;/ APPLICATION NUMBER: PCT/US95/10194
;/ FILING DATE:
;/ CLASSIFICATION:
;/ ATTORNEY/AGENT INFORMATION:
;/ NAME: White, John P.
;/ REGISTRATION NUMBER: 28,678
;/ REFERENCE/DOCKET NUMBER: 45185-C-PCT/JPW/MS
;/ TELECOMMUNICATION INFORMATION:
;/ TELEPHONE: (212) 278-0400
;/ TELEFAX: (212) 391-0525
;/ INFORMATION FOR SEQ ID NO: 20:
;/ SEQUENCE CHARACTERISTICS:
;/ LENGTH: 2259 base pairs
;/ TYPE: nucleic acid
;/ STRANDEDNESS: single
;/ TOPOLOGY: linear
;/ MOLECULE TYPE: DNA (genomic)
;/ HYPOTHETICAL: N
;/ ANTI-SENSE: N
;/ FEATURE:
;/ NAME/KEY: CDS
;/ LOCATION: 1..2259
;/ OTHER INFORMATION:
;/ PCT-US95-10194-20

Query Match 1.0%; Score 18; DB 5; Length 2259;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 396 TCTGGCAGCTCTACAGT 413
Db 1194 TCTGGCAGCTCTACAGT 1211

RESULT 11
US-09-085-199B-44
;/ Sequence 44, Application US/09085199B
;/ Patent No. 6235879
;/ GENERAL INFORMATION:
;/ APPLICANT: Hayden, Michael R.
;/ APPLICANT: Hackam, Abigail
;/ APPLICANT: Hug, A.H.M. Mahbubul
;/ APPLICANT: Chopra, Vikramjit Singh
;/ APPLICANT: Kalchman, Michael
;/ TITLE OF INVENTION: Apoptosis Modulators That Interact with the
;/ TITLE OF INVENTION: Huntington's Disease Gene
;/ NUMBER OF SEQUENCES: 44
;/ CORRESPONDENCE ADDRESS:
;/ ADDRESSEE: Oppedahl & Larson
;/ STREET: PO Box 5270
;/ CITY: Frisco
;/ STATE: CO
;/ COUNTRY: USA
;/ ZIP: 80443-5270
;/ COMPUTER READABLE FORM:
;/ MEDIUM TYPE: Diskette, 3.50 inch, 1.44 Kb storage
;/ COMPUTER: IBM Compatible
;/ OPERATING SYSTEM: MS DOS 5.0
;/ SOFTWARE: WordPerfect
;/ CURRENT APPLICATION DATA:
;/ APPLICATION NUMBER: US/09/085,199B
;/ FILING DATE:
;/ CLASSIFICATION:
;/ ATTORNEY/AGENT INFORMATION:
;/ NAME: Larson, Marina T.
;/ REGISTRATION NUMBER: 32038
;/ REFERENCE/DOCKET NUMBER: UBC.P-013US2
;/ TELECOMMUNICATION INFORMATION:
;/ TELEPHONE: (970) 668-2052
;/ TELEFAX: (970) 668-2052
;/ INFORMATION FOR SEQ ID NO: 44:
;/ SEQUENCE CHARACTERISTICS:

;/ LENGTH: 3715
;/ TYPE: nucleic acid
;/ STRANDEDNESS: double
;/ TOPOLOGY: linear
;/ MOLECULE TYPE: genomic DNA
;/ HYPOTHETICAL: no
;/ ANTI-SENSE: no
;/ ORIGINAL SOURCE: human
;/ ORGANISM: human
;/ FEATURE:
;/ OTHER INFORMATION: exon 29 and partial cds of HIP1
;/ US-09-085-199B-44

Query Match 1.0%; Score 18; DB 4; Length 3715;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 252 CTCTGATCTGGGAAA 269
Db 1724 CTCTGATCTGGGAAA 1741

RESULT 12
US-09-085-199B-3
;/ Sequence 3, Application US/09085199B
;/ Patent No. 6235879
;/ GENERAL INFORMATION:
;/ APPLICANT: Hayden, Michael R.
;/ APPLICANT: Hackam, Abigail
;/ APPLICANT: Hug, A.H.M. Mahbubul
;/ APPLICANT: Chopra, Vikramjit Singh
;/ APPLICANT: Kalchman, Michael
;/ TITLE OF INVENTION: Apoptosis Modulators That Interact with the
;/ TITLE OF INVENTION: Huntington's Disease Gene
;/ NUMBER OF SEQUENCES: 44
;/ CORRESPONDENCE ADDRESS:
;/ ADDRESSEE: Oppedahl & Larson
;/ STREET: PO Box 5270
;/ CITY: Frisco
;/ STATE: CO
;/ COUNTRY: USA
;/ ZIP: 80443-5270
;/ COMPUTER READABLE FORM:
;/ MEDIUM TYPE: Diskette, 3.50 inch, 1.44 Kb storage
;/ COMPUTER: IBM Compatible
;/ OPERATING SYSTEM: MS DOS 5.0
;/ SOFTWARE: WordPerfect
;/ CURRENT APPLICATION DATA:
;/ APPLICATION NUMBER: US/09/085,199B
;/ FILING DATE:
;/ CLASSIFICATION:
;/ ATTORNEY/AGENT INFORMATION:
;/ NAME: Larson, Marina T.
;/ REGISTRATION NUMBER: 32038
;/ REFERENCE/DOCKET NUMBER: UBC.P-013US2
;/ TELECOMMUNICATION INFORMATION:
;/ TELEPHONE: (970) 668-2052
;/ TELEFAX: (970) 668-2052
;/ INFORMATION FOR SEQ ID NO: 3:
;/ SEQUENCE CHARACTERISTICS:
;/ LENGTH: 4796
;/ TYPE: nucleic acid
;/ STRANDEDNESS: single
;/ TOPOLOGY: linear
;/ MOLECULE TYPE: cDNA
;/ HYPOTHETICAL: no
;/ ANTI-SENSE: no
;/ ORIGINAL SOURCE: human
;/ ORGANISM: human
;/ FEATURE:
;/ OTHER INFORMATION: cDNA for Huntington-interacting protein
;/ US-09-085-199B-3

Query Match 1.0%; Score 18; DB 4; Length 4796;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 252 CTTCTGATCTTGGGAAA 269
Db 4717 CTTCTGATCTTGGGAAA 4734

RESULT 13
US-08-420-235B-1/c
Sequence 1, Application US/08420235B
Patent No. 5801042
GENERAL INFORMATION:
APPLICANT: Chang, Yuan
APPLICANT: Moore, Patrick S.
TITLE OF INVENTION: UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS
TITLE OF INVENTION: SEQUENCES AND USES THEREOF
NUMBER OF SEQUENCES: 47
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/420,235B
FILING DATE:
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 45185-B
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 278-0400
TELEFAX: (212) 391-0525
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 20710 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: N
ANTI-SENSE: N
US-08-420-235B-1

Query Match 1.0%; Score 18; DB 1; Length 20710;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 396 TCTGGGACGCTCTACAGT 413
Db 6565 TCTGGGACGCTCTACAGT 5548

RESULT 14
US-08-793-624-1/c
Sequence 1, Application US/08793624C
Patent No. 6150093
GENERAL INFORMATION:
APPLICANT: Chang, Yuan
APPLICANT: Moore, Patrick S.
TITLE OF INVENTION: Unique Associated Kaposi's Sarcoma Virus Sequences And
TITLE OF INVENTION: Uses Thereof
FILE REFERENCE: 45185-C-PCT-US/JPW
CURRENT APPLICATION NUMBER: US/08/793,624C

CURRENT FILING DATE: 1997-02-18
NUMBER OF SEQ ID NOS: 58
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 1
LENGTH: 20710
TYPE: DNA
ORGANISM: Kaposi's sarcoma-associated herpesvirus
US-08-793-624-1

Query Match 1.0%; Score 18; DB 3; Length 20710;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 396 TCTGGGACGCTCTACAGT 413
Db 6565 TCTGGGACGCTCTACAGT 6548

RESULT 15
PCT-US95-10194-1/c
Sequence 1, Application PC/TUS9510194
GENERAL INFORMATION:
APPLICANT: The Trustees of Columbia University in the City of New York
APPLICANT: City
TITLE OF INVENTION: UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS
TITLE OF INVENTION: SEQUENCES AND USES THEREOF
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/10194
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 45185-C-PCT/JPW/MSC
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 278-0400
TELEFAX: (212) 391-0525
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 20710 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: N
ANTI-SENSE: N
PCT-US95-10194-1

Query Match 1.0%; Score 18; DB 5; Length 20710;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 396 TCTGGGACGCTCTACAGT 413
Db 6565 TCTGGGACGCTCTACAGT 6548

RESULT 16
US-08-961-527-83
Sequence 83, Application US/08961527

```
; Patent No. 6420135
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 391
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,527
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 83:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 28473 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-08-961-527-83

Query Match 1.0%; Score 18; DB 4; Length 28473;
Best Local Similarity 100.0%; Pred.No.22;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 133 GAAAAATCATGCCAGTGT 150
Db 17357 GAAAAATCATGCCAGTGT 17374

RESULT 17
US-08-770-379-18/c
; Sequence 18, Application US/08770379
; Patent No. 5849564
; GENERAL INFORMATION:
; APPLICANT: Chang, Yuan
; APPLICANT: Bohenzky, Roy A.
; APPLICANT: Russo, James J.
; APPLICANT: Edelman, Isidore S.
; APPLICANT: Moore, Patrick S.
; TITLE OF INVENTION: POLYPEPTIDES FROM KAPOSI'S SARCOMA-ASSOCIATED
; HERPESVIRUS, DNA ENCODING SAME AND USES THEREOF
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/757,669A
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 45185-F
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 35100 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-757-669A-18
```

```
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/770,379
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 52342
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 35100 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-770-379-18

Query Match 1.0%; Score 18; DB 2; Length 35100;
Best Local Similarity 100.0%; Pred.No.22;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 396 TCTGGCAGCTCTACAGT 413
Db 6485 TCTGGCAGCTCTACAGT 6468

RESULT 18
US-08-757-669A-18/c
; Sequence 18, Application US/08757669A
; Patent No. 6183751
; GENERAL INFORMATION:
; APPLICANT: Chang, Yuan
; APPLICANT: Bohenzky, Roy A.
; APPLICANT: Russo, James J.
; APPLICANT: Edelman, Isidore S.
; APPLICANT: Moore, Patrick S.
; TITLE OF INVENTION: UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS
; SEQUENCES AND USES THEREOF
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/757,669A
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 45185-F
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 35100 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-757-669A-18
```

Query Match 1.0%; Score 18; DB 4; Length 35100;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 396 TCTGGGAGCTCTACAGT 413
Db 6485 TCTGGGAGCTCTACAGT 6468

RESULT 19
US-09-230-371A-18/c
; Sequence 18, Application US/09230371A
; Patent No. 6348586

; GENERAL INFORMATION:
; APPLICANT: Chang, Yuan
; APPLICANT: Bohenzky, Roy A.
; APPLICANT: Russo, James J
; APPLICANT: Edelman, Tsidore S
; APPLICANT: Moore, Patrick S
; TITLE OF INVENTION: UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS SEQUENCES AND
; FILE REFERENCE: 45185-G-PCT-US
; CURRENT APPLICATION NUMBER: US/09/230,371A
; CURRENT FILING DATE: 1999-11-17
; PRIOR APPLICATION NUMBER: PCT/US97/13346
; PRIOR FILING DATE: 1997-07-22
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 18
; LENGTH: 35100
; TYPE: DNA
; ORGANISM: Kaposi's sarcoma-associated herpesvirus
US-09-230-371A-18

Query Match 1.0%; Score 18; DB 4; Length 35100;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 396 TCTGGGAGCTCTACAGT 413
Db 6485 TCTGGGAGCTCTACAGT 6468

RESULT 20
US-09-749-588-3/c
; Sequence 3, Application US/09749588
; Patent No. 6423521

; GENERAL INFORMATION:
; APPLICANT: CHANDRAMOULISWARAN, Ishwar et al
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CU001068
; CURRENT APPLICATION NUMBER: US/09/749,588
; CURRENT FILING DATE: 2000-12-28
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 36159
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)-(36159)
; OTHER INFORMATION: n = A,T,C or G
US-09-749-588-3

Query Match 1.0%; Score 18; DB 4; Length 36159;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 829 TGTTCTGGAGTCTTT 846

Db 27845 TGTTCTGGAGTCTTT 27828

RESULT 21
US-09-090-793-12
; Sequence 12, Application US/09090793
; Patent No. 6140486

; GENERAL INFORMATION:
; APPLICANT: Calgene, LLC
; TITLE OF INVENTION: Production of polyunsaturated fatty acids by expression
; TITLE OF INVENTION: of polypeptide-like synthesis genes in plants
; FILE REFERENCE: CGN 131.01US
; CURRENT APPLICATION NUMBER: US/09/090,793
; CURRENT FILING DATE: 1998-06-04
; EARLIER APPLICATION NUMBER: 60/048,650
; EARLIER FILING DATE: 1997-06-04
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12
; LENGTH: 40138
; TYPE: DNA
; ORGANISM: Vibrio marinus
US-09-090-793-12

Query Match 1.0%; Score 18; DB 3; Length 40138;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 324 TTTTGATGAGCGTATATT 341
Db 551 TTTTGATGAGCGTATATT 568

RESULT 22
US-08-756-771-2/c
; Sequence 2, Application US/08756771
; Patent No. 5817497

; GENERAL INFORMATION:
; APPLICANT: Goll, Surya K.
; APPLICANT: Hillman, Jennifer L.
; TITLE OF INVENTION: A NOVEL GLUTATHIONE S-TRANSFERASE
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: US
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/756,771
; FILING DATE: Herewith
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0162 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 800 base pairs

```
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: Consensus
; CLONE: Consensus
US-08-756-771-2

Query Match          0.9%; Score 17; DB 1; Length 800;
Best Local Similarity 100.0%; Pred. No. 61;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 674 TTAGGAATTTCTCTTC 690
Db 518 TTAGGAATTTCTCTTC 502

RESULT 23
US-09-596-571-2/c
; Sequence 2, Application US/09096571
; Patent No. 5976528
; GENERAL INFORMATION:
; APPLICANT: Goli, Surya K.
; TITLE OF INVENTION: A NOVEL GLUTATHIONE S-TRANSFERASE
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: US
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/096,571
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/096,571
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0162 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 800 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: Consensus
; CLONE: Consensus
US-09-309-320-2

Query Match          0.9%; Score 17; DB 2; Length 800;
Best Local Similarity 100.0%; Pred. No. 61;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 674 TTAGGAATTTCTCTTC 690
Db 518 TTAGGAATTTCTCTTC 502

US-09-096-571-2
; Sequence 2, Application US/09096571
; Patent No. 5976528
; GENERAL INFORMATION:
; APPLICANT: Goli, Surya K.
; TITLE OF INVENTION: A NOVEL GLUTATHIONE S-TRANSFERASE
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: US
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/096,571
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/756,771
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0162 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 800 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: Consensus
; CLONE: Consensus
US-09-096-571-2

Query Match          0.9%; Score 17; DB 2; Length 800;
Best Local Similarity 100.0%; Pred. No. 61;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 674 TTAGGAATTTCTCTTC 690
Db 518 TTAGGAATTTCTCTTC 502
```

```
RESULT 24
US-09-309-320-2/c
; Sequence 2, Application US/09309320
; Patent No. 6248325
; GENERAL INFORMATION:
; APPLICANT: Goli, Surya K.
; APPLICANT: Hillman, Jennifer L.
; TITLE OF INVENTION: A NOVEL GLUTATHIONE S-TRANSFERASE
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: US
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/309,320
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/096,571
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0162 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 800 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: Consensus
; CLONE: Consensus
US-09-309-320-2

Query Match          0.9%; Score 17; DB 4; Length 800;
Best Local Similarity 100.0%; Pred. No. 61;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 674 TTAGGAATTTCTCTTC 690
Db 518 TTAGGAATTTCTCTTC 502

RESULT 25
US-08-617-8608-21
; Sequence 21, Application US/086178608
; Patent No. 6133506
; GENERAL INFORMATION:
; APPLICANT: Typer, R., Bautor, J., Bothmann, H., Filsak, E.,
; APPLICANT: Hvricker-Grandpierre, C., Klein, B., Martini, N.,
; APPLICANT: Mller, A., Schulte, W., Voetz, M., Walek, J.,
; APPLICANT: Schell, J.
; TITLE OF INVENTION: Promoters
; NUMBER OF SEQUENCES: 35
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Steinberg, Raskin & Davidson, P.C.
; STREET: 1140 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
```

ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4 Mb storage
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25 (PFO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/617,860B
FILING DATE: 01-MAR-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/EP94/02950
FILING DATE: 05-SEP-1994
APPLICATION NUMBER: DE P4329951.2
FILING DATE: 04-SEP-1993
INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 926 base pairs
TYPE: Nucleic acid
STRANDEDNESS: Double stranded
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Cuphea lanceolata
IMMEDIATE SOURCE:
LIBRARY: genomic Lambda FIX II
CLONE: C1KRg3
FEATURE:
NAME/KEY: TATA-Signal
LOCATION: 827..838
FEATURE:
NAME/KEY: Transcription start
LOCATION: 864
FEATURE:
NAME/KEY: Startcodon
LOCATION: 916..918
FEATURE:
NAME/KEY: CDS
LOCATION: 916..926
US-08-617-860B-21

Query Match 0.9%; Score 17; DB 3; Length 926;
Best Local Similarity 100.0%; Pred. No. 61;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 216 CTTTGTCCCGAGTT 232
DB 459 CTTTGTCCCGAGTT 475

RESULT 26
US-09-045-583-6
Sequence 6, Application US/09045583
Patent No. 6287805
GENERAL INFORMATION:
APPLICANT: Graham, Gerard J. et al.
TITLE OF INVENTION: No. 6287805el Molecules of the G Protein-Coupled
NUMBER OF SEQUENCES: 56
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHYE & COCKFIELD, LLP
STREET: 28 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/045,583

FILING DATE: 20-MAR-98
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Mandragouras, Amy E.
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: MNI-044
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)742-4214
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 1137 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1134
US-09-045-583-6

Query Match 0.9%; Score 17; DB 4; Length 1137;
Best Local Similarity 100.0%; Pred. No. 62;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 720 TTTTCTTCATGCTGTGG 736
DB 770 TTTTCTTCATGCTGTGG 786

RESULT 27
US-09-534-185-6
Sequence 6, Application US/09534185
Patent No. 6403767
GENERAL INFORMATION:
APPLICANT: Graham, Gerard J. et al.
TITLE OF INVENTION: No. 6403767el Molecules of the G Protein-Coupled
Heptahelical Receptor Superfamily and Uses
Therefor
NUMBER OF SEQUENCES: 56
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHYE & COCKFIELD, LLP
STREET: 28 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/534,185
FILING DATE: 24-Mar-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/045,583
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Mandragouras, Amy E.
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: MNI-044
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)742-4214
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 1137 base pairs
TYPE: nucleic acid

NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 79
LENGTH: 1692
TYPE: DNA
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-79

Query Match 0.9%; Score 17; DB 4; Length 1692;
Best Local Similarity 100.0%; Pred. No. 63;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1210 TTGCTCAAGGTAGCA 1226
Db 799 TTGCTCAAGGTAGCA 815

RESULT 31
US-08-936-165A-116/c

Sequence 116, Application US/08936165A
Patent No. 6348582

GENERAL INFORMATION:

APPLICANT: Black, Michael

APPLICANT: Burnham, Martin

APPLICANT: Hodgson, John

APPLICANT: Knowles, David

APPLICANT: Lometto, Michael

APPLICANT: Nicholas, Richard

APPLICANT: Pratt, Julie

APPLICANT: Reichard, Richard

APPLICANT: Rosenberg, Martin

APPLICANT: Ward, Judith

TITLE OF INVENTION: No. 6348582el Prokaryotic Polynucleotides,

TITLE OF INVENTION: Polypeptides and Their Uses

NUMBER OF SEQUENCES: 534

CORRESPONDENCE ADDRESS:

ADDRESSEE: SmithKline Beecham Corporation

STREET: 709 Swedeland Road

CITY: King of Prussia

STATE: PA

COUNTRY: USA

ZIP: 19406-0939

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/936,165A

FILING DATE: 24-SEP-1997

CLASSIFICATION: 536

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/027,032

FILING DATE: 24-SEP-1996

ATTORNEY/AGENT INFORMATION:

NAME: Giuntini, Edward R

REGISTRATION NUMBER: 38,891

REFERENCE/DOCKET NUMBER: P50549

TELECOMMUNICATION INFORMATION:

TELEPHONE: 610-270-4478

TELEFAX: 610-270-5090

TELEX:

INFORMATION FOR SEQ ID NO: 116:

SEQUENCE CHARACTERISTICS:

LENGTH: 1789 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: Genomic DNA

US-08-936-165A-116

Query Match 0.9%; Score 17; DB 4; Length 1789;
Best Local Similarity 100.0%; Pred. No. 63;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 189 CCTCAGATTTAAATCT 205
Db 94 CCTCAGATTTAAATCT 78

RESULT 32

US-09-484-970B-165

Sequence 165, Application US/09484970B

Patent No. 6426186

GENERAL INFORMATION:

APPLICANT: Jones, Karen A.

APPLICANT: Volkmuth, Wayne

APPLICANT: Walker, Michael G

TITLE OF INVENTION: BONE REMODELING GENES

FILE REFERENCE: PB-0014 US

CURRENT APPLICATION NUMBER: US/09/484,970B

CURRENT FILING DATE: 2000-01-18

NUMBER OF SEQ ID NOS: 172

SOFTWARE: PERL Program

SEQ ID NO 165

LENGTH: 2114

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: misc feature

OTHER INFORMATION: Incyte ID No. 6426186 199021.1

NAME/KEY: unsure

LOCATION: 1907, 1909, 2009, 2014

OTHER INFORMATION: a, t, c, g, or other

US-09-484-970B-165

Query Match 0.9%; Score 17; DB 4; Length 2114;
Best Local Similarity 100.0%; Pred. No. 63;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 537 TTAATAATCTGTAGAA 553
Db 1449 TTAATAATCTGTAGAA 1465

RESULT 33

US-09-066-047-17/c

Sequence 17, Application US/09066047A

Patent No. 6306394

GENERAL INFORMATION:

APPLICANT: MURPHY, Cheryl

STOREY, James

BEUTZ, Gerald A.

COUGHLIN, Richard T.

TITLE OF INVENTION: NUCLEIC ACIDS, PROTEINS, AND METHODS OF

USE OF GRANULOCYTIC ERLHICHA

NUMBER OF SEQUENCES: 41

CORRESPONDENCE ADDRESS:

ADDRESSEE: HALE AND DORR LLP

STREET: 60 State Street

CITY: Boston

STATE: Massachusetts

COUNTRY: United States

ZIP: 02109

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/066,047A

FILING DATE: 24-Apr-1998

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 60/044,869

FILING DATE: 25-APR-1997

ATTORNEY/AGENT INFORMATION:

US-09-066-047-17/c

NAME: Superko, Colleen
REGISTRATION NUMBER: 39,850
REFERENCE/DOCKET NUMBER: 106,941,156
TELEPHONE: (617) 526-6000
TELEFAX: (617) 526-5000
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 2900 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
SEQUENCE DESCRIPTION: SEQ ID NO: 17:
US-09-066-047-17

Query Match 0.9%; Score 17; DB 4; Length 2900;
Best Local Similarity 100.0%; Pred. No. 64;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 192 CAGATTAAATCTGGA 208
Db 1674 CAGATTAAATCTGGA 1658

RESULT 34
US-08-720-484A-1/c
Sequence 1, Application US/08720484A
Patent No. 5990281
GENERAL INFORMATION:
APPLICANT: Desauvage, Frederic
APPLICANT: Rosenthal, Arnon
APPLICANT: Stone, Donna
TITLE OF INVENTION: Vertebrate Smoothened Proteins
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 1 DNA Way
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080

COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WinPatIn (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/720,484A
FILING DATE: 30-Sep-1996
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Svoboda, Craig G.
REGISTRATION NUMBER: 39,044
REFERENCE/DOCKET NUMBER: P1050
TELEPHONE: 650/225-1489
TELEFAX: 650/952-9881
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3854 base pairs
TYPE: Nucleic Acid
STRANDEDNESS: Single
TOPOLOGY: Linear
US-08-720-484A-1

Query Match 0.9%; Score 17; DB 2; Length 3854;
Best Local Similarity 100.0%; Pred. No. 65;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1280 TTGGCAATCATCTTGCT 1296

Db 2173 TTGGCAATCATCTTGCT 2157
RESULT 35
US-08-953-823A-1/c
Sequence 1, Application US/08953823A
Patent No. 6136958
GENERAL INFORMATION:
APPLICANT: Desauvage, Frederic
APPLICANT: Rosenthal, Arnon
APPLICANT: Stone, Donna
TITLE OF INVENTION: Vertebrate Smoothened Proteins
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 1 DNA Way
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080

COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WinPatIn (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/953,823A
FILING DATE: 30-Sep-1997
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/027070
FILING DATE: 30-Sep-1996
ATTORNEY/AGENT INFORMATION:
NAME: Svoboda, Craig G.
REGISTRATION NUMBER: 39,044
REFERENCE/DOCKET NUMBER: P1050R1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650/225-1489
TELEFAX: 650/952-9881
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3854 base pairs
TYPE: Nucleic Acid
STRANDEDNESS: Single
TOPOLOGY: Linear
US-08-953-823A-1

Query Match 0.9%; Score 17; DB 3; Length 3854;
Best Local Similarity 100.0%; Pred. No. 65;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1280 TTGGCAATCATCTTGCT 1296
Db 2173 TTGGCAATCATCTTGCT 2157

RESULT 36
US-09-398-239-1/c
Sequence 1, Application US/09398239
Patent No. 6407216
GENERAL INFORMATION:
APPLICANT: Desauvage, Frederic
APPLICANT: Rosenthal, Arnon
APPLICANT: Stone, Donna
TITLE OF INVENTION: Vertebrate Smoothened Proteins
FILE REFERENCE: P1050D1
CURRENT APPLICATION NUMBER: US/09/398,239
CURRENT FILING DATE: 1999-09-20
EARLIER APPLICATION NUMBER: US 08/720,484
EARLIER FILING DATE: 1996-09-30
NUMBER OF SEQ ID NOS: 5
SEQ ID NO 1

LENGTH: 3854
TYPE: DNA
ORGANISM: Rattus No. 6407216vegicus
US-09-398-239-1

Query Match 0.9%; Score 17; DB 4; Length 3854;
Best Local Similarity 100.0%; Pred. No. 65;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1280 TTGGCAATCATCTTGT 1296
Db 2173 TTGGCAATCATCTTGT 2157

RESULT 37
US-09-221-017B-8/c
Sequence 8, Application US/09221017B
Patent No. 6444799
GENERAL INFORMATION:
APPLICANT: Ross, Bruce C.
TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF
NUMBER OF SEQUENCES: 1120
CORRESPONDENCE ADDRESSES:
ADDRESSEE: MORRISON & FOERSTER
STREET: 755 PAGE MILL ROAD
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304-1018
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows
SOFTWARE: FastSeq for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/221,017B
FILING DATE: 23-DEC-1998
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PFI182
FILING DATE: 31-DEC-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PFI546
FILING DATE: 30-JAN-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PFI2911
FILING DATE: 09-APR-1998
APPLICATION DATA:
APPLICATION NUMBER: PFI/AU98/01023
FILING DATE: 10-DEC-1998
ATTORNEY/AGENT INFORMATION:
NAME: Montoy, Gladys H.
REGISTRATION NUMBER: 32,430
REFERENCE/DOCKET NUMBER: 27340-20021.00
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-813-5600
TELEFAX: 650-494-0792
TELEX: 706141
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 6089 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: UNKNOWN
ORIGINAL SOURCE:
ORGANISM: PORYPHYROMONAS GINGIVALIS
FEATURE:
NAME/KEY: misc feature
LOCATION: 1...6089
US-09-221-017B-8

Query Match 0.9%; Score 17; DB 4; Length 6089;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 234 TCAGTGAGAGAGATC 250
Db 4247 TCAGTGAGAGAGATC 4231

RESULT 38
US-08-975-762-46/c
Sequence 46, Application US/08975762
Patent No. 6207169
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Lodes, Michael J.
APPLICANT: Houghton, Raymond
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS AND
NUMBER OF SEQUENCES: 73
CORRESPONDENCE ADDRESSES:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/975,762
FILING DATE: 21-MAR-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.439
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-682-4900
TELEFAX: 206-682-6031
INFORMATION FOR SEQ ID NO: 46:
SEQUENCE CHARACTERISTICS:
LENGTH: 7091 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-975-762-46

Query Match 0.9%; Score 17; DB 4; Length 7091;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1781 CGGTGCAATATATAA 1797
Db 4451 CGGTGCAATATATAA 4435

RESULT 39
US-09-295-028-46/c
Sequence 46, Application US/09295028
Patent No. 6277381
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Lodes, Michael J.
APPLICANT: Houghton, Raymond L.
APPLICANT: McNeill, Patricia D.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS
AND TREATMENT OF EHRILCHIA INFECTION
FILE REFERENCE: 210121.439C4

```
; CURRENT APPLICATION NUMBER: US/09/295,028
; CURRENT FILING DATE: 1999-04-20
; NUMBER OF SEQ ID NOS: 85
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 46
; LENGTH: 7091
; TYPE: DNA
; ORGANISM: Ehrlichia sp.
US-09-295-028-46

Query Match          0.9%; Score 17; DB 4; Length 7091;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1781 CGGTAGCAATATAAAA 1797
|||||
Db 4451 CGGTAGCAATATAAAA 4435

RESULT 40
US-09-106-582-46/c
; Sequence 46, Application US/09106582
; Patent No. 6306402
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS AND
; THERAPY
; NUMBER OF SEQUENCES: 73
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/106,582
; FILING DATE: 29-JUN-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.439C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-622-4900
; TELEFAX: 206-682-6031
; INFORMATION FOR SEQ ID NO: 46:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7091 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-09-106-582-46

Query Match          0.9%; Score 17; DB 4; Length 7091;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1781 CGGTAGCAATATAAAA 1797
|||||
Db 4451 CGGTAGCAATATAAAA 4435

RESULT 41
US-08-961-527-189/c
; Sequence 189, Application US/08961527
```

```
; Patent No. 6420135
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 391
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,527
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 189:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7174 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-961-527-189

Query Match          0.9%; Score 17; DB 4; Length 7174;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 531 AAAGTCTTAAAAATCT 547
|||||
Db 2311 AAAGTCTTAAAAATCT 2295

RESULT 42
US-08-961-527-167
; Sequence 167, Application US/08961527
; Patent No. 6420135
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 391
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,527
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
```

APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Andrews
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 167:
SEQUENCE CHARACTERISTICS:
LENGTH: 9711 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-961-527-167

Query Match 0.9%; Score 17; DB 4; Length 9711;
Best Local Similarity 100.0%; Pred. No. 68;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1762 ACCGTAACCTATCT 1778
Db 7007 ACCGTAACCTATCT 7023

RESULT 43
US-08-038-768A-4/c
Sequence 4, Application US/08038768A
Patent No. 5569828
GENERAL INFORMATION:
APPLICANT: McMillen, Michael D.; Roth, Bradley A.; Townsend,
APPLICANT: Rod
TITLE OF INVENTION: MAIZE CHLOROTIC DWARF VIRUS RESISTANCE
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pioneer Hi-Bred International, Inc.
STREET: 700 Capital Square, 400 Locust
CITY: Des Moines
STATE: Iowa
COUNTRY: United States
ZIP: 50309
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS, Microsoft Windows
SOFTWARE: Microsoft Windows No. 5569828epad
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/038.768A
FILING DATE: 19930324
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Roth, Michael J.
REGISTRATION NUMBER: 29,342
REFERENCE/DOCKET NUMBER: 0235 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (515) 245-3594
TELEFAX: (515) 245-3634
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 11785 bases
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: viral RNA
HYPOTHETICAL: NO
US-08-038-768A-4

Query Match 0.9%; Score 17; DB 1; Length 11785;

Best Local Similarity 100.0%; Pred. No. 69;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1736 CGTAAGAAATCGCTC 1752
Db 965 CGTAAGAAATCGCTC 949

RESULT 44
US-08-416-603-3/c
Sequence 3, Application US/08416603
Patent No. 5866780
GENERAL INFORMATION:
APPLICANT: Law, Marcus
APPLICANT: Hebata, Uedare
APPLICANT: Reddick, Bradford B.
TITLE OF INVENTION: Maize Chlorotic Dwarf Virus Genome and
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Saliwanchik & Saliwanchik
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: FL
COUNTRY: USA
ZIP: 32606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30B
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/416,603
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Lloyd, Jeffrey
REGISTRATION NUMBER: 35,589
TELECOMMUNICATION INFORMATION:
TELEPHONE: 904-375-8100
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 11785 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
HYPOTHETICAL: NO
FEATURE:
NAME/KEY: misc feature
LOCATION: 1..11785
OTHER INFORMATION: /note="CDNA sequence of MCDV-T
OTHER INFORMATION: genome as disclosed in WO 94/21796"
FEATURE:
NAME/KEY: CDS
LOCATION: 456..10826
US-08-416-603-3

Query Match 0.9%; Score 17; DB 2; Length 11785;
Best Local Similarity 100.0%; Pred. No. 69;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1736 CGTAAGAAATCGCTC 1752
Db 965 CGTAAGAAATCGCTC 949

RESULT 45
US-08-687-080-105/c
Sequence 105, Application US/08687080
Patent No. 5965427
GENERAL INFORMATION:
APPLICANT: Gregory Dolganov

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 17, 2002, 08:59:21 ; Search time 3371 Seconds
(without alignments)
16092.444 Million cell updates/sec

Title: US-09-868-987-1
Perfect score: 1864
Sequence: 1 atggacttcgcattgtc.....atgacaaattcagataatgc 1864

Scoring table: OLIGO NUC
Gapop 60.0 , Gapext 60.0

Searched: 2054640 seqs, 14551402878 residues

Word size : 0
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : GenEmbl.*

- 1: gb.ba.*
- 2: gb.htg.*
- 3: gb.in.*
- 4: gb.om.*
- 5: gb.ov.*
- 6: gb.pat.*
- 7: gb.ph.*
- 8: gb.pl.*
- 9: gb.pr.*
- 10: gb.ro.*
- 11: gb.sts.*
- 12: gb.sy.*
- 13: gb.un.*
- 14: gb.vi.*
- 15: em.ba.*
- 16: em.fun.*
- 17: em.hum.*
- 18: em.in.*
- 19: em.mu.*
- 20: em.om.*
- 21: em.or.*
- 22: em.ov.*
- 23: em.pat.*
- 24: em.ph.*
- 25: em.pl.*
- 26: em.ro.*
- 27: em.sts.*
- 28: em.un.*
- 29: em.vi.*
- 30: em.htg.hum.*
- 31: em.htg.inv.*
- 32: em.htg.other.*
- 33: em.htg.mus.*
- 34: em.htg.pln.*
- 35: em.htg.rod.*
- 36: em.htg.mam.*
- 37: em.htg.vrt.*
- 38: em.ey.*
- 39: em.htgo.hum.*
- 40: em.htgo.mus.*
- 41: em.htgo.other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	1453	78.0	14134	1	AE001641	AE001641 Chlamydia
C 2	1453	78.0	300550	1	AP002547	AP002547 Chlamydia
C 3	1421	76.2	12198	1	AE002179	AE002179 Chlamydia
C 4	92	4.9	11938	1	AE002180	AE002180 Chlamydia
C 5	22	1.2	3183	5	AF019977	AF019977 Gallus ga
C 6	22	1.2	151032	9	AC105201	AC105201 Homo sapi
C 7	22	1.2	159198	2	AC124363	AC124363 Mus muscu
C 8	22	1.2	172579	9	AC036125	AC036125 Homo sapi
C 9	22	1.2	209636	2	AC121084	AC121084 Mus muscu
C 10	22	1.2	247074	2	AC121990	AC121990 Mus muscu
C 11	22	1.2	293272	2	AC120685	AC120685 Rattus no
C 12	21	1.1	15389	1	AF026147	AF026147 Bacillus
C 13	21	1.1	41830	3	CBRG44E08	AC084613 Caenorhab
C 14	21	1.1	90157	9	AL355552	AL355552 Human DNA
C 15	21	1.1	94960	9	AC000066	AC000066 Homo sapi
C 16	21	1.1	100170	9	AL161743	AL161743 Human DNA
C 17	21	1.1	110000	2	AC122235_2	Continuation (3 of
C 18	21	1.1	114741	2	AC124954	AC124954 Medicago
C 19	21	1.1	123091	2	AC129991	AC129991 Rattus no
C 20	21	1.1	124028	2	AC137017	AC127017 Rattus no
C 21	21	1.1	133054	2	AL445200	AL445200 Homo sapi
C 22	21	1.1	145661	2	AC114333	AC114333 Canis fam
C 23	21	1.1	147803	9	AL356867	AL356867 Human DNA
C 24	21	1.1	157326	2	AC011024	AC011024 Homo sapi
C 25	21	1.1	168632	2	AC106409	AC106409 Rattus no
C 26	21	1.1	170008	9	HS010770	AJ010770 Homo sapi
C 27	21	1.1	175809	2	AC096466	AC096466 Rattus no
C 28	21	1.1	207730	1	BSUB0011	239114 Bacillus su
C 29	21	1.1	207890	2	AL732291	AL732291 Mus muscu
C 30	21	1.1	219622	9	AC009630	AC009630 Homo sapi
C 31	21	1.1	257967	2	AL365337	AL365337 Mus muscu
C 32	20	1.1	1472	3	AF322045	AF322045 Malawimon
C 33	20	1.1	2895	1	CAU38234	U38234 Clostridium
C 34	20	1.1	3644	3	IGU48714	U48714 Idiomyia gr
C 35	20	1.1	7370	5	AF043383	AF043383 Pleuroneoc
C 36	20	1.1	9989	1	AE001318	AE001318 Chlamydia
C 37	20	1.1	10742	1	AE007683	AE007683 Clostridi
C 38	20	1.1	10836	1	AE014001	AE014001 Yersinia
C 39	20	1.1	12442	1	AE002341	AE002341 Chlamydia
C 40	20	1.1	42632	3	CEP52H3	Z66512 Caenorhabdi
C 41	20	1.1	72371	2	AC119990	AC119990 Mus muscu
C 42	20	1.1	73256	2	AC069119	AC069119 Homo sapi
C 43	20	1.1	74105	2	AC087648	AC087648 Homo sapi
C 44	20	1.1	84574	9	AL713969	AL713969 Human DNA
C 45	20	1.1	101730	2	AC111865	AC111865 Rattus no
C 46	20	1.1	102086	9	AL512447	AL512447 Human DNA
C 47	20	1.1	102846	9	AC002483	AC002483 Homo sapi
C 48	20	1.1	111357	10	AC110422	AC110422 Mus muscu
C 49	20	1.1	113425	2	AP004256	AP004256 Oryza sat
C 50	20	1.1	127334	9	AC109579	AC109579 Homo sapi
C 51	20	1.1	133072	9	HS1097P24	AL035258 Human DNA
C 52	20	1.1	134703	2	AC124259	AC124259 Homo sapi
C 53	20	1.1	135373	2	AP004695	AP004695 Oryza sat
C 54	20	1.1	135378	9	AL390060	AL390060 Human DNA
C 55	20	1.1	138824	9	AC008689	AC008689 Homo sapi
C 56	20	1.1	141267	2	AC013694	AC013694 Homo sapi
C 57	20	1.1	142100	2	AC094067	AC094067 Rattus no
C 58	20	1.1	142600	9	AC005371	AC005371 Homo sapi
C 59	20	1.1	143024	2	AC098435	AC098435 Rattus no
C 60	20	1.1	146528	2	AC106923	AC106923 Rattus no
C 61	20	1.1	147009	2	AC084179	AC084179 Homo sapi
C 62	20	1.1	150204	8	OSJN00136	AL662934 Oryza sat
C 63	20	1.1	154761	2	AP005312	AP005312 Oryza sat
C 64	20	1.1	159791	9	AC007179	AC007179 Homo sapi
C 65	20	1.1	160799	2	AC021601	AC021601 Homo sapi

c 942	18	1.0	167052	2	AC128655	Homo sapi
c 943	18	1.0	167093	2	AC094951	Rattus no
944	18	1.0	167164	2	AC026506	Homo sapi
945	18	1.0	167231	2	AC115367	Rattus no
946	18	1.0	167302	2	AC099756	Homo sapi
c 947	18	1.0	167398	9	CNS07EET	AL445886 Human chr
c 948	18	1.0	167525	9	AC006531	Homo sapi
c 949	18	1.0	167568	2	AC111734	Rattus no
c 950	18	1.0	167749	2	AC073142	Homo sapi
c 951	18	1.0	167753	9	AC108752	Homo sapi
c 952	18	1.0	167833	2	AC022749	Homo sapi
c 953	18	1.0	167863	2	AC121949	Mus muscu
c 954	18	1.0	168056	2	AC004458	Homo sapi
c 955	18	1.0	168066	2	AC026902	Homo sapi
c 956	18	1.0	168097	2	AC102401	Mus muscu
c 957	18	1.0	168253	2	AC116818	Mus muscu
c 958	18	1.0	168319	2	AC023282	Homo sapi
c 959	18	1.0	168332	2	AC053483	Homo sapi
c 960	18	1.0	168357	9	AC104440	Homo sapi
c 961	18	1.0	168433	2	AC110696	Rattus no
c 962	18	1.0	168478	2	AL611925	Homo sapi
c 963	18	1.0	168483	2	AC094611	Rattus no
c 964	18	1.0	168505	2	AC080010	Homo sapi
c 965	18	1.0	168644	9	AC004544	Homo sapi
c 966	18	1.0	168677	9	CNS0000F	AL049837 Human chr
c 967	18	1.0	168818	2	AC013496	Homo sapi
c 968	18	1.0	168822	2	AC007511	Homo sapi
c 969	18	1.0	168946	2	AC125639	Rattus no
c 970	18	1.0	168975	2	AC122511	Mus muscu
c 971	18	1.0	169030	2	AC125625	Homo sapi
c 972	18	1.0	169106	2	AC120743	Rattus no
c 973	18	1.0	169229	2	AC125913	Rattus no
c 974	18	1.0	169237	9	AC009516	Homo sapi
c 975	18	1.0	169248	2	AC129338	Homo sapi
c 976	18	1.0	169408	9	AC079258	Homo sapi
c 977	18	1.0	169537	2	AC107329	Rattus no
c 978	18	1.0	169542	2	AL356482	Homo sapi
c 979	18	1.0	169572	2	AC023832	Mus muscu
c 980	18	1.0	169607	2	AC023569	Homo sapi
c 981	18	1.0	169660	9	AL161449	Human DNA
c 982	18	1.0	169670	9	AL330962	Human DNA
c 983	18	1.0	169814	2	AC130759	Rattus no
c 984	18	1.0	169877	3	AC104612	Drosophi
c 985	18	1.0	170008	9	AC117391	Homo sapi
c 986	18	1.0	170145	2	AC113969	Mus muscu
c 987	18	1.0	170215	9	AC092691	Homo sapi
c 988	18	1.0	170241	2	AC120635	Rattus no
c 989	18	1.0	170251	2	AC009935	Homo sapi
c 990	18	1.0	170291	2	AC106944	Rattus no
c 991	18	1.0	170316	2	AC122151	Macaca mu
c 992	18	1.0	170332	2	AC106104	Rattus no
c 993	18	1.0	170393	9	AC026473	Homo sapi
c 994	18	1.0	170414	2	AC026019	Homo sapi
c 995	18	1.0	170472	2	AC125911	Rattus no
c 996	18	1.0	170512	2	AC021806	Homo sapi
c 997	18	1.0	170578	2	AC110703	Rattus no
c 998	18	1.0	170608	2	AL359372	Homo sapi
c 999	18	1.0	170749	2	AC112890	Rattus no
c1000	18	1.0	170772	2	AC014654	Homo sapi

ALIGNMENTS

RESULT 1
LOCUS AE001641 14134 bp DNA linear BCT 01-Dec-2000
DEFINITION Chlamydia pneumoniae section 57 of 103 of the complete genome.
ACCESSION AE001641 AE001363
VERSION AE001641.1 GI:4376855

KEYWORDS Chlamydia pneumoniae CWL029.
SOURCE Chlamydia pneumoniae
ORGANISM Chlamydia pneumoniae CWL029

REFERENCE
1 (bases 1 to 14134)
Kallan, S., Mitchell, W., Marathe, R., Lamell, C., Fan, J., Hyman, R.W., Olinger, L., Grimwood, J., Davis, R.W. and Stephens, R.S.
Comparative genomes of Chlamydia pneumoniae and C. trachomatis
Nat. Genet. 21 (4), 385-389 (1999)

TITLE
JOURNAL
MEDLINE
99206606

PUBMED
10192388

REFERENCE
2 (bases 1 to 14134)
Kallan, S., Mitchell, W., Marathe, R., Lamell, C., Fan, J., Olinger, L., Grimwood, J., Davis, R.W. and Stephens, R.S.
Direct Submission
Submitted (01-DEC-1998) Program in Infectious Diseases, University of California, 235 Earl Warren Hall, Berkeley, CA 94720, USA
Location/Qualifiers
1. 14134
/organism="Chlamydia pneumoniae CWL029"
/strain="CWL029"
/db_xref="taxon:115713"
/complement(158, .1924)
/gene="recJ"
/note="Cpn0563"
complement(158, .1924)
/gene="recJ"
/codon_start=1
/transl_table=1
/product="sedDNA Exonuclease"
/protein_id="AA018703.1"
/db_xref="GI:4376856"
/translation="MTNSDNASAGLILMHPKEDPAFGIMTIKEPHLPPTVAOIFISR
GROTOEIHKEFLYSLSYDGLDQSKSKEATERRLADREHEMHTGSDSDCMG
VALVEPLRIDVHSYIFPLGAIKQHGKSTSLIKKAREGTTLLITDCCITAGKEV
SDITQIGDVIITDHPMPGKIPHCVATLNPGLRHTYPNRELTGVGAFLKARGLN
ALISNLVPSQSGSLKLLDVTLLGTDVGLDGNPMVYGIKEIARGAPGLNK
LALCGVSEKSEVTSTDIYLKAPKINSGLRDLDPKAGVELLTDODERDALIMELDN
INRERORIEAEVODVOETLNPPEILNOAALVLSSTMAHARVIFISARLAKTYNKP
VITIAIORIGIGKSARITGSFPLGLVKCSLLISXGHDPAAGVAVGVDEPKK
KRVHLNSSLKKKGDPLPHIEDIAYDFAIDIDLLASHELFPFGKGNLMPFIYSKVR
OVRYPKVLPGNHLKLYLSQKEMNLGVAFGLGRHADLAKASWHPLEIAYTRLSQTS
GSGVHLVLRDPRISSEPRFSD"
complement(2010, .6218)
/gene="secD/secF"
/note="Cpn0564"
complement(2010, .6218)
/gene="secD/secF"
/codon_start=1
/transl_table=1
/product="Protein Export Proteins SecD/SecF (fusion)"
/protein_id="AA018704.1"
/db_xref="GI:4376857"

CDS
1
/transl_table=1
/product="Protein Export Proteins SecD/SecF (fusion)"
/protein_id="AA018704.1"
/db_xref="GI:4376857"

gene
1
/transl_table=1
/product="Protein Export Proteins SecD/SecF (fusion)"
/protein_id="AA018704.1"
/db_xref="GI:4376857"

```
gene
CDS
/translation="MIFRICKFTWAFSLPYKLVYGVKKNFIKGPAAIIVAHNNSPL
DPIALHMCVHECIYHLARSLFNIPLWLMKQWCFPPRODEGNSAAFKIARSLFNKRKK
LVYIPEGARSPDQOLQGVKIGMAKSRVPIIPVIRGTFFAHRHOKIPHVMTKI
TCVFGTMYFDDIIQNPENIKNKETYQIITNTQNMKIAELKAWVESGCKGVP"
10975..12660
/gene="args"
/note="Cpno570"
10975..12660
/gene="args"
/codon_start=1
/transl_table=11
/product="Arginyl tRNA Transferase"
/protein_id="AADI8710.1"
/db_xref="GI:4376863"
/translation="NSTLSILSVCSQAIKAFPNLEDWAPETITSTKEHFGHYQCN
ALMKLQVQSPKKIIDFSPNIAKMHVGHRLSTIIGDSLARIIFSVMHDLVRLNHI
GDMGTAFGLMITYLOENPCDYSLEDLTSLYKAYVCFNDEBEFKKESQONVALQAK
DPOAIWEKICETSEKAFQKVIDLDIIVVEKKGESFYNPFLEIIEDELEKGLLTVS
NDAKCVHEAFSIPFMVQSDGYNATTDLAAMRYRIEBEDHADKIIIVTDLQOSLHF
QLLEATAIAAGYLQPGIFSHVGFGLVDLPQKKLKTSGENVKRLLEDLTAIEKAEAA
LREHPELIDEAIOERAPVIGINAIKYSDLSHRTSDYVSEFKMLRFEGNTAMFLLY
AVVRIQIKRRLGISOLSEGPPEIOEPAEELIALTLRPEALESTIKELCPHFLTD
YLYNLTHKNGFPERDSHIQDSPYAKSRLFLCALAEQVLATGMHLGLKTLERL"
complement(12614..13990)
/gene="murA"
/note="Cpno571"
complement(12614..13990)
/gene="murA"
/codon_start=1
/transl_table=11
/product="UDP-N-Acetylglucosamine Transferase"
/protein_id="AADI8711.1"
/db_xref="GI:4376864"
/translation="MOIAOVFCGRNGEVKUSGAKNAATKLLVASILSDOKCTLRNV
PDIGCVSLVELCKSLCAHVSMDKETEVELEIYPTQCTRVPTFSNVRNIPILLIGA
LIGRCPEGVPTVGGDAIGERTLNFHFGELKOLGVOISDSSGYSYAKAPRGLKNGVI
HLPVSGVATENILIAAIHAKRTVINKVALEAIBLDLVLQAGADITTDNDRTID
IFGTGGLSGVDHTILPDKIEAASFMAAVSGGVFRVNAKQELLIPFLKMLRSIGGG
FLVSESGIEFFQERPLVGVVLETDVHPGFLTDWQPPFVALLSQAGSSVIHETVHEN
RLGYLHLOHMGAECCOLFHQCLSTKACVAINGFPHSAVTHGATPLWASHLVIDLPA
GFAYVMAALIAEAGGSGSIIENTHLLDRGYTNWVGLSLGAKIKIOPDMEQBELITSFKS
LAURASL"
BASE COUNT 4180 a 2981 c 2805 g 4168 t
ORIGIN
Query Match 78.0%; Score 1453; DB 1; Length 14134;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1863; Conservative 0; Mismatches 1; Indels 3; Gaps 3;
QY 1 ATGGACTTCGGCATATTGTTCAGAGGGGATCAGCGGCACCTGCTAATGGACAATATTCCTGC 60
DB 3771 ATGGACTTCGGCATATTGTTCAGAGGGGATCAGCGGCACCTGCTAATGGACAATATTCCTGC 3712
QY 61 AAACCGTGGATGGCGTATGCTGTAGTGATTGACGGTTATATGCTCAGCAGCCCTATTTT 120
DB 3711 AAACCGTGGATGGCGTATGCTGTAGTGATTGACGGTTATATGCTCAGCAGCCCTATTTT 3652
QY 121 AAACGTCCTCCATGGAAATCATGCCAGTGCTCAGGGAATTTTACCACCGTGAAGTGAG 180
DB 3651 AAACGTCCTCCATGGAAATCATGCCAGTGCTCAGGGAATTTTACCACCGTGAAGTGAG 3592
QY 181 CAAACTCGCTCAGATTAAATACTCGAGCGATGCTCTTTTGTTCGCCAGGTTCTCAGTGA 240
DB 3591 CAAACTCGCTCAGATTAAATACTCGAGCGATGCTCTTTTGTTCGCCAGGTTCTCAGTGA 3532
QY 241 AGAGACGATCTCTTCTGATCTTGGGAAAAAACAATGTACAAAGGCATTAATCTCAGCATG 300
DB 3531 AGAGACGATCTCTTCTGATCTTGGGAAAAAACAATGTACAAAGGCATTAATCTCAGCATG 3472
QY 301 CTGTGGCTTGGCAATGCTTATTTGTTGATGAGCGTATATATAGATTGAGGCGTCAT 360
DB 3471 CTGTGGCTTGGCAATGCTTATTTGTTGATGAGCGTATATATAGATTGAGGCGTCAT 3412
```

```
gene
CDS
/translation="MIFRICKFTWAFSLPYKLVYGVKKNFIKGPAAIIVAHNNSPL
DPIALHMCVHECIYHLARSLFNIPLWLMKQWCFPPRODEGNSAAFKIARSLFNKRKK
LVYIPEGARSPDQOLQGVKIGMAKSRVPIIPVIRGTFFAHRHOKIPHVMTKI
TCVFGTMYFDDIIQNPENIKNKETYQIITNTQNMKIAELKAWVESGCKGVP"
10975..12660
/gene="args"
/note="Cpno570"
10975..12660
/gene="args"
/codon_start=1
/transl_table=11
/product="CT449 hypothetical protein"
/protein_id="AADI8705.1"
/db_xref="GI:4376858"
/translation="MKFPRISDLIPTQMWWRGGNVHVPNAONLKKILGVLV
ACFGLALLGCAAPAAAGVQOTIFPCIGILMILGLVLLGPAYLQYSGWGRFPRPLRETK
VFEKPIWLGCLSLQSKWKIRPCGYHPGCPQVEICEGQEIWTKIFQKSRDNTSI
FLQEMQIALRQIESKSSURKTFADPSVSSLLSEIOREEQYIDPKVIMSSSD
QASDTHPKAIYVNIISDAQEPQRCYIDATYKAFETVLDDQIDPNIVKKHTIYLT
PIGVDPALPKEQENIKLASQAFLYSAEQVAKRMREKQDSIRIKFIFTDPTSPS
LYSPHHSYTPHSVTPISLSGFGVEQESYTPA"
8009..8761
/gene="yaeS"
/note="Cpno566"
8009..8761
/gene="yaeS"
/codon_start=1
/transl_table=11
/product="yaeS family hypothetical protein"
/protein_id="AADI8706.1"
/db_xref="GI:4376859"
/translation="MSLATNNAKSPFLQRLPNHVAIIMDGNRRWYKXHEGCHTH
TSGHYGAICVLPNINAVLDLGIKVLTYTFSTENFCRKEEIOEINIFVTDLDKOL
PYLMENECILKICGLDKSLPKGIOTKINHVRMTASPSRLVLAVNYSKDKELVRAP
KLUHVDLNLKISDDSESLSISYLTSLTSDLPDLPDIRTIGGEMRVNLFMLQIAYTE
LYITDLWDPDTPODLFEAINVYQSRRRGK"
8762..9688
/gene="cdsA"
/note="Cpno567"
8762..9688
/gene="cdsA"
/codon_start=1
/transl_table=11
/product="Phosphatidate Cytidylyltransferase"
/protein_id="AADI8707.1"
/db_xref="GI:4376860"
/translation="MLNSKFKSGTAYGDLFQRVVHSLVLFVLLVLLVSSLPPLTS
FALGFIATCGAVGTYSMAKAMHYPLSTFSAIGSFLFALSFLSIRWHSLPFGF
FDALPWTLLVWVMSIFVRKSTIGALQSGVTLFSLYVGPILRFLHVLVSFIHT
QEPVLGWASFLIATKGADIFGYFFGKAFGNKKIAPOISPNKTVGVFVAGCLGATL
ISPIFQIPIPTRASYPMPAILIPLGLALGITGFFGDIIESIFKRD AHLKNSNKLKA
VGMGLDPLDSLLSTPIAYLFLTLTQSKETIG"
9689..10335
/gene="cmk"
/note="Cpno568"
9689..10335
/gene="cmk"
/codon_start=1
/transl_table=11
/product="CMP kinase"
/protein_id="AADI8708.1"
/db_xref="GI:4376861"
/translation="MITIDGSGTGSTAKALADHLHFNQNTGMYRTLAYRLQ
SPWATPLTLFLEPPSPFTATGQPLESFNGHLITSELTTOEVANAASELSQLEPV
RAFMDQIRRYAQLGNCVFFGRDMGKVPFNADLKIFLTSSPEVRAQRRLKDPGL
SPQOLQABLRYDAADAQRAHPLVIPENGIVIVDSSDLTIROVLEKILALLFRNEL"
10332..10970
/gene="plcC"
/note="Cpno569"
10332..10970
/gene="plcC"
/codon_start=1
/transl_table=11
/product="Glycerol-3-P Acyltransferase"
/protein_id="AADI8709.1"
/db_xref="GI:4376862"
```

QY 361 CGCTTGGGAGCTGTTCTTGTGAATCTTTGCTTATCTGGCAGCTCTACAGTATTGGA 420
 DB 3411 CGCTTGGGAGCTGTTCTTGTGAATCTTTGCTTATCTGGCAGCTCTACAGTATTGGA 3352
 QY 421 TGGGCCATCAGCTGTCAGAGACTGCTGGGATGTTCTTGCTATGAGGAGATGGCCGTAGA 480
 DB 3351 TGGGCCATCAGCTGTCAGAGACTGCTGGGATGTTCTTGCTATGAGGAGATGGCCGTAGA 3292
 QY 481 TGGCAATGTTCTTGATTTCCGAAAGATCCGAGAGAAATTTTATTTGTCTCAAGTCTTAA 540
 DB 3291 TGGCAATGTTCTTGATTTCCGAAAGATCCGAGAGAAATTTTATTTGTCTCAAGTCTTAA 3222
 QY 541 AAAATCTGTAAGAAAAGATATACCAAGGCTTTTGAGGCCATTTTGTATCTTAATTTGAC 600
 DB 3231 AAAATCTGTAAGAAAAGATATACCAAGGCTTTTGAGGCCATTTTGTATCTTAATTTGAC 3172
 QY 601 TACAGTATGAGCCCTCAGACACTTCTTCTCTAGATACAGGGGCTATTAAGGGTTTGC 660
 DB 3171 TACAGTATGAGCCCTCAGACACTTCTTCTCTAGATACAGGGGCTATTAAGGGTTTGC 3112
 QY 661 TTTGACATGATTTTATGAAATTTTCTCTCAATGTTTACGGCTCTTTTCAATGACTAAAT 720
 DB 3111 TTTGACATGATTTTATGAAATTTTCTCTCAATGTTTACGGCTCTTTTCAATGACTAAAT 3052
 QY 721 TTTCTCATGCTGTGATGATATAGAACCCCAATACACAGTTGATGATGATTAAGTT 780
 DB 3051 TTTCTCATGCTGTGATGATATAGAACCCCAATACACAGTTGATGATGATTAAGTT 2992
 QY 781 CGTGGGATTAAGATATTTCTTGAGAGATGCAAAAACTTGGGCTGTTTCTGAGAG 840
 DB 2991 CGTGGGATTAAGATATTTCTTGAGAGATGCAAAAACTTGGGCTGTTTCTGAGAG 2932
 QY 841 TGTATTTCTTTAGAGTTGCTTGTCTCTCGAGTTTGAGACCTGGAATTCGGTTTGGGAAT 900
 DB 2931 TGTATTTCTTTAGAGTTGCTTGTCTCTCGAGTTTGAGACCTGGAATTCGGTTTGGGAAT 2872
 QY 901 GGATTTTAAAGAGAGGATATGCTTAACTTAACTCAAAAGAGATGAGCATCAGCATGT 960
 DB 2871 GGATTTTAAAGAGAGGATATGCTTAACTTAACTCAAAAGAGATGAGCATCAGCATGT 2812
 QY 961 TGTCTCAATGCTGTGCAAACTGTGCAATTAACAGGAGCTGCTCTTCTTAGAGA 1020
 DB 2811 TGTCTCAATGCTGTGCAAACTGTGCAATTAACAGGAGCTGCTCTTCTTAGAGA 2752
 QY 1021 CTTCGATTTCAAACTTTGATCTTCGAAAGATCAAAATCTATTTTATGATTAAGC 1080
 DB 2751 CTTCGATTTCAAACTTTGATCTTCGAAAGATCAAAATCTATTTTATGATTAAGC 2692
 QY 1081 TTTAAGCTATACT-AAGCAGATACGAGCTCTCTCTAAATTAACGATCAGTACGTGC 1139
 DB 2691 TTTAAGCTATACT-AAGCAGATACGAGCTCTCTCTAAATTAACGATCAGTACGTGC 2632
 QY 1140 GTTA-TTGTGGGATTTGTTGTCAGAAACAGGCTAGATTTCTCTAGGAAATCTTAACGA 1198
 DB 2631 GTTA-TTGTGGGATTTGTTGTCAGAAACAGGCTAGATTTCTCTAGGAAATCTTAACGA 2572
 QY 1199 AAGCAAAA-TTTTGTCAAGGTAAGCAGCAAACTATCGAAGAAAATGCTTATCAGGC 1257
 DB 2571 AAGCAAAAATTTTGTCAAGGTAAGCAGCAAACTATCGAAGAAAATGCTTATCAGGC 2512
 QY 1258 GACCATCGGGCTTTAGAGACCTTGGCAATCATCTGCTATATGAGTTTGGCCTTTGA 1317
 DB 2511 GACCATCGGGCTTTAGAGACCTTGGCAATCATCTGCTATATGAGTTTGGCCTTTGA 2452
 QY 1318 ATGCGAATATGCTTTCAGTCCGATATGCGCTTAAATCATACCTTTTGGCTACTGTGC 1377
 DB 2451 ATGCGAATATGCTTTCAGTCCGATATGCGCTTAAATCATACCTTTTGGCTACTGTGC 2392
 QY 1378 AGTCTGTTTATGACACTTCTTTTGAAGAAAATCAATATGATTTGCAAGCATTTGG 1437
 DB 2391 AGTCTGTTTATGACACTTCTTTTGAAGAAAATCAATATGATTTGCAAGCATTTGG 2332

QY 1438 TGGTTAATGACGTAATGAGGGGATTCATTAACAATACTTGTGATGATTTGATGCTAT 1497
 DB 2331 TGGTTAATGACGTAATGAGGGGATTCATTAACAATACTTGTGATGATTTGATGCTAT 2272
 QY 1498 TCGTGAAGATCGCCAGACCAACCTGTTTACCCCTATGATGATGTTTATGATGATCCCT 1557
 DB 2271 TCGTGAAGATCGCCAGACCAACCTGTTTACCCCTATGATGATGTTTATGATGATCCCT 2212
 QY 1558 TCAAAAGAGCTTACCGCCGACGGTAATGACACAGCTACATCTATCAGTTTGTAAAT 1617
 DB 2211 TCAAAAGAGCTTACCGCCGACGGTAATGACACAGCTACATCTATCAGTTTGTAAAT 2152
 QY 1618 GCTTTGTTATATGAGGGGCTCTCTGCTTAAATTTTGAATTAATATGACCAATAGGAT 1677
 DB 2151 GCTTTGTTATATGAGGGGCTCTCTGCTTAAATTTTGAATTAATATGACCAATAGGAT 2092
 QY 1678 TCTTCTAGGAATCTTATCGCTCTTATATGACACCACTCTGTTTATGATGCTCG 1737
 DB 2091 TCTTCTAGGAATCTTATCGCTCTTATATGACACCACTCTGTTTATGATGCTCG 2032
 QY 1738 TAAAGAAATCGCTCAAAATTAAGTACCGTTAACTTAATCTAAAGTATGACCAATATATA 1797
 DB 2031 TAAAGAAATCGCTCAAAATTAAGTACCGTTAACTTAATCTAAAGTATGACCAATATATA 1972
 QY 1798 ATCTCTTTGGGACTTATGACCAAGGCCCTGTGATTAATTTATGACCAATATGAC 1857
 DB 1971 ATCTCTTTGGGACTTATGACCAAGGCCCTGTGATTAATTTATGACCAATATGAC 1912
 QY 1858 ATATATGC 1864
 DB 1911 ATATATGC 1905

RESULT 2
 AP002547/c
 LOCUS 300550 bp DNA linear BCF 25-MAY-2002
 DEFINITION Chlamydia pneumoniae J138 genomic DNA, complete sequence,
 section 3/4.
 ACCESSION AP002547 AB033786 AB033816 AB033817 BA000008
 VERSION AP002547.2 GI:10176693
 KEYWORDS
 SOURCE Chlamydia pneumoniae J138 (strain:J138) DNA.
 ORGANISM Chlamydia pneumoniae J138
 Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.

REFERENCE
 AUTHORS
 1 Shirai, M., Hirakawa, H., Ouchi, K., Tabuchi, M., Kishi, F., Kimoto, M., Takeuchi, A., Nishida, J., Shibata, K., Fujinaga, R., Yoneda, H., Matsushima, H., Tanaka, C., Furukawa, S., Miura, K., Nakazawa, A., Ishii, K., Shiba, T., Hattori, M., Kohara, S. and Nakazawa, T.
 Comparison of outer membrane protein genes omp and pmp in the whole genome sequences of Chlamydia pneumoniae isolates from Japan and the United States
 J. Infect. Dis. 181 Suppl 3, S524-S527 (2000)
 20238986

JOURNAL
 MEDLINE
 REFERENCE
 AUTHORS
 2 Shirai, M., Hirakawa, H., Kimoto, M., Tabuchi, M., Kishi, F., Ouchi, K., Shiba, T., Ishii, K., Hattori, M., Kohara, S. and Nakazawa, T.
 Comparison of whole genome sequences of Chlamydia pneumoniae J138 from Japan and CML029 from USA
 Nucleic Acids Res. 28 (12), 2311-2314 (2000)
 20330349

JOURNAL
 MEDLINE
 REFERENCE
 AUTHORS
 3 (bases 1 to 300550)
 Shirai, M.
 Direct Submission
 Submitted (04-JUL-2000) Mutsunori Shirai, Yamaguchi University
 School of Medicine, Department of Microbiology, 1-1-1
 Minamikogushi, Ube, Yamaguchi 755-8505, Japan
 (E-mail: mshirai@po.cc.yamaguchi-u.ac.jp, Tel: 81-836-22-2227,
 Fax: 81-836-22-2415)
 On or before Sep 15, 2000 this sequence version replaced
 gi:6172298, gi:6172300, gi:6172396, gi:6172398, gi:8978889.
 COMMENT
 FEATURES
 source
 1. 300550


```

/organism="Chlamydomophila pneumoniae J138"
/strain="J138"
/db_xref="taxon:138677"
/notes="synonym:Chlamydia pneumoniae (strain J138)"
complement(21..794)
/gene="dapP"
complement(21..794)
/gene="dapP"
/codon_start=1
/transl_table=11
/product="diaminopimelate epimerase"
/protein_id="BAA98725.1"
/db_xref="GI:8978890"
/translation="MAYFSPSTISKYFIYSGAGNRFLLGETLPEVDVRLCQTRVD
GFLYKPSSCADAGLIIFNSDGSPTWCGNLRCAIAHLASQKGSKISVSTDSGLYS
GFLYSWRVLVDMTLADRAVSHRESRDPDPKPVVCIHTGVPHAVIILPEISTLDL
SILFGVPTLRYHOTFSPDGVNWNFVQILGHCOLRVTYERGVEGETAACGTGALASLV
SNSVGKESIOIHTWGEMLTVSQNRGVLYQSVTRDL"
complement(763..1338)
/gene="clpP_1"
complement(763..1338)
/gene="clpP_1"
/codon_start=1
/transl_table=11
/product="Clp protease"
/protein_id="BAA98726.1"
/db_xref="GI:8978891"
/translation="MADGEVHKLRDIIIEKELLEARRVFFSPVTEKSASDAIKKLWYL
ELKDPGPIVFINPSGSDVAGFAVMDIOIKMLTSPVTVTVGLAASMGSVLSICAAP
GRRFATPSRIMIHQPSIGGPIITCGAOTDLIHAREILKTKARIIDVVEATNQPRDI
EKALDRNMWTANAKDFGLDGLIFSNFL"
complement(1358..2917)
/gene="glyA"
complement(1358..2917)
/gene="glyA"
/codon_start=1
/transl_table=11
/product="serine hydroxymethyltransferase"
/protein_id="BAA98727.1"
/db_xref="GI:8978892"
/translation="MLKVFKEFKFAIVEITKVAVVUSLLHKLLENASGKGKOSLAS
TAYLAALDHLNAPFSIGERIIDLKQSRSHLKMIASENSSLSVLQMGNLITDKYC
EGSPFKPFGYSCCENDAIENECVETAKELFADCAQVPHSGQGLNLAVALIUTHKV
QGAVALKGYKTVNELTEETLLKAEKSCVCLGPSLNSGGHLTHGNVRLNWSKLM
RCFYDNPAGTDECDDYAEISRLAKYKPKVLIAGYSYRRLNFAVLKQIAEDCGSVL
WVDAHFGVLVAGGVFVDEENPIYADIVTTTHKTLRGGGLVLAATREYESTLKA
CPLMMGQPLPHVIAAKTVALKEALSVDPKYAHQVNNARLAEFLSHGLRLITGCT
DNHMVVIDLGSGLGSKIAEDILSSVGIANNLSLPSDAIGKWDTSGLRGTPTALTF
QMGIDEEVADIIKVLNRILRSCHVEGSSKQKQELDPEAIAQARDVRNLLLRFP
LYEIDIEALV"
3107..3778
/gene="CPj0522"
3107..3778
/gene="CPj0522"
/codon_start=1
/transl_table=11
/product="CT433 hypothetical protein"
/protein_id="BAA98728.1"
/db_xref="GI:8978893"
/translation="MTLYLGLNQTKARKYQAHYLPILTLPPYAKSTQNKALQFLQ
ATHVILSPSSTHFLFSRWTSLLSKATLKTLYICIGESTKERLLSLFGQKYVYATQ
EIAHGIPPLQALPSSARILYPHSSLARPIREFLYRFTFSPHYTVPRKLUKNI
LSKYKKIIFTSPTVRFAKIFPRFPKTYWCQGRMTLQEFQKFSQKQVSLLETGK
SRTSP"
3843..4175
/gene="CPj0523"
3843..4175
/gene="CPj0523"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAA98729.1"
/db_xref="GI:8978894"

```

```

/translation="MASSATPGPDGTAPSLPPATPRPNFKLAFVTIATIALWIAL
IATTIAGLCIHPLCSFILTALPLYFISYICSHYARNVYIALDVVVDHSHKLODMRS
HSPISFSDR"
4223..5302
/gene="CPj0524"
4223..5302
/gene="CPj0524"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAA98730.1"
/db_xref="GI:8978895"
/translation="MSGPRTSSQSVLSVYPRDKELIAPKQFTIAKISTLAILASL
ALGALVAGISLTIVLGNPVLALLITFALPSVTVFLVYHOMTSKVSNNWOKVLEQNFK
PLGAKWQKVCYCNEMQYNNHNPFKVAIOYDASQFPQFTFLGLRVIEKNQST
GIIFNPVPTNLINDTANTSLTILYSLDKVSDCKQREGGPAKEDGFPSPTEVRV
VKLPNALDQTNLNLSSAKKSIILPTFLGHVGPKEELPNOQEVYRQALLAVENCL
KAAIESHAALVALPLETSVYVPEEILPKEGTFYWDNQTAQCKRALLDAIQNTALR
YPORSLLVLQDPFNIESQSRSEE"
5642..6406
/gene="CPj0525"
5642..6406
/gene="CPj0525"
/codon_start=1
/transl_table=11
/product="CT398 hypothetical protein"
/protein_id="BAA98731.1"
/db_xref="GI:8978896"
/translation="MDALLSIIAIOELDIKMLRMVKEHOKELAKVOSLKSDIRR
KVQEKLEMLNLTQJRDCENRIOEISEQINKLENOQAAVKMDEFNALQEMTANK
ERRSELQSLMDKQAGGDLIVLSKLASTENSSVIEKEIFESIKINEEGKAL
LEQRTBLKHATNPELLISYIERLLNNKKDVRVFIENRVCSGCHIVLTPQHENVLRKID
RLIFCBHCSRILYQESQVNAQENAKRRRRRAAV"
complement(6827..7816)
/gene="kpsF"
complement(6827..7816)
/gene="kpsF"
/codon_start=1
/transl_table=11
/product="GutQ/kpsF family sugar-P isomerase"
/protein_id="BAA98732.1"
/db_xref="GI:8978897"
/translation="MFSPISTDVCDILKQKEAVDFFQAFQPKRAMQLAEKILGH
SGVFFSGVSGKCVAKRLVATLQSLSERALFFSPVDLLHGLSLGSPGDIVCLFSKS
TCQWILDFTVPHLKSRAILVAITSMPSYLNLAALSDLVLPVSLAEIDPFLNIPNTST
TCQWIFDFTLMLLFHSRGSVLSYTKGNHPSGQVKANGKVDKFMFPKTEVPFCHLG
DKVFSLEVSAYCGCCVCIVDQFLMGIFTDGLRSLASYGGEVLSLSLEKVMTA
NPRCIEDSDIALALQLMESSSPVAVLPVLIDNEENRHVTGLLHHTLAKAGLL"
complement(7843..9024)
/gene="sucB_2"
complement(7843..9024)
/gene="sucB_2"
/codon_start=1
/transl_table=11
/product="dihydrolipoamide succinyltransferase"
/protein_id="BAA98733.1"
/db_xref="GI:8978898"
/translation="MIFEPFPKIGETSSGGSIVRLKNGLDHVARDEPLIEVSTDKI
ATLPSPKAGRLVFCVNEGDEVASGDVLGLIELEBISEADDESTCPLTSCETKSEA
SSSSSVWFSPAVLSLAOREGIDNLQKTAGCKGGRVTRQDLIAVISRSQVSIPE
IFQGVNRIISPLRLAASSLSKSDSEVPHSLVVDVTDLNLNLSGSRQRFLDTH
GVKLITTSFTVQCLAQTLRQFPLNGSLDGTIVMKKSNNVGVAVNLNKEGVVFPVTH
NCQDRGLIVSKALADLSSARLNKLDPPSEVQSGSVTVTFGMTGALIGMPIIRYPEV
AILGIGTIQKRVVVRDDDSLAIKRVVYVTLTFDHRVLDGIYSGSEFLTSLKNRLESVTM
G"
complement(9038..10282)
/gene="gltr"
complement(9038..10282)
/gene="gltr"
/codon_start=1
/transl_table=11
/product="glutamate symport"
/protein_id="BAA98734.1"

```

```

/db_xref="GI:8978899"
/translation="MKLMKIFIFGVYTLGLVLEDKAIFPKPIDIEPLNLSMNV
PLVFCSMVIGIASISDMKLGRIKSVLYLGTALALVIGLCRAMVSPENGDPA
QASMDSATVIDSNKTAAYFLJIAQVPSNPVRSFAGNILOJIIIFIPJIALRL
SEBERGPVERFIDFSEIMLRWNMIMSFPYGVNEMAMISGNHGLVLMQDKEFI
AYLICTFARLIVFGGLVAFSGCKMSFKLSMDMDISCAVSPSSALPLPYMBCVS
KMLGVSAEVSFVPLPLGATVNNNGRIIFQGMVAVPIAQNVPDISLSLLILVYATF
SAVGSAGVPGGMITLGSVLASVLPIDGIALIILAGIDRLDIVGIPMILGDVAVATI
VASGEBELSPYESIKQESVEIT"
gene
  complement(10282..11379)
  /gene="ycaH"
  complement(10282..11379)
  /gene="ycaH"
  /codon_start=1
  /transl_table=11
  /product="Atpase"

Query Match      78.3%; Score 1453; DB 1; Length 300550;
Best Local Similarity 99.3%; Pred. No. 0;
Matches 1863; Conservative 0; Mismatches 1; Indels 3; Gaps 3;

QY 1 ATGACATTCGCATATTTGTAGAGAGGGATCAGCGCATCTAATGAGCAATATTTCTGC 60
Db 51024 ATGACATTCGCATATTTGTAGAGAGGGATCAGCGCATCTAATGAGCAATATTTCTGC 50965

QY 61 AAACCGTGATGGCGATGCTGTAGTATGATGACGCTTATATGCTCAGACCCCTATTTT 120
Db 50964 AAACCGTGATGGCGATGCTGTAGTATGATGACGCTTATATGCTCAGACCCCTATTTT 50905

QY 121 AAACGTCCATTTGAAAATATGTCAGTGTCTCAGGAAATTTTACCAACCGTGAAAGAG 180
Db 50904 AAACGTCCATTTGAAAATATGTCAGTGTCTCAGGAAATTTTACCAACCGTGAAAGAG 50845

QY 181 CAAATCGCCTCAGATTTTAAATCTGAGAGCATGCTTTGTTCCGAGGTTCTCAGTGA 240
Db 50844 CAAATCGCCTCAGATTTTAAATCTGAGAGCATGCTTTGTTCCGAGGTTCTCAGTGA 50785

QY 241 AGAGACGATCTCTTCTGATTTGGGAAAAAACAATGACACAGCATTTATCTCAGCATG 300
Db 50784 AGAGACGATCTCTTCTGATTTGGGAAAAAACAATGACACAGCATTTATCTCAGCATG 50725

QY 301 CTGTGGCTTGGCAATGCTTATTTGTTTGAAGGCTATATGATTTGGAGGCGTCAT 360
Db 50724 CTGTGGCTTGGCAATGCTTATTTGTTTGAAGGCTATATGATTTGGAGGCGTCAT 50665

QY 361 CGCTTGGGAGCTGTCTTGTGAATCTTTGCTTATCTGGGACCTCTACAGTATTTGA 420
Db 50664 CGCTTGGGAGCTGTCTTGTGAATCTTTGCTTATCTGGGACCTCTACAGTATTTGA 50605

QY 421 TGGCCACATCACTTGTCAAGACTCGCTGGGATGTTCTTCTATAGGGATGCGCGTGA 480
Db 50604 TGGCCACATCACTTGTCAAGACTCGCTGGGATGTTCTTCTATAGGGATGCGCGTGA 50545

QY 481 TGCAAATGTTTGTATTTCAAGAAATCCGAGAGAAATTTTATGTCCTCAAGCTTAA 540
Db 50544 TGCAAATGTTTGTATTTCAAGAAATCCGAGAGAAATTTTATGTCCTCAAGCTTAA 50485

QY 541 AAAATCTGTAGAAAAGATATACCAAGCTTTTGGAGCCATTTTGTATCTTAACCTGAC 600
Db 50484 AAAATCTGTAGAAAAGATATACCAAGCTTTTGGAGCCATTTTGTATCTTAACCTGAC 50425

QY 601 TACAGTATGGCTCAGCAGCTTTTCTTCTAGATACAGGGCTATTTAAAGGGTTTGC 660
Db 50424 TACAGTATGGCTCAGCAGCTTTTCTTCTAGATACAGGGCTATTTAAAGGGTTTGC 50365

QY 661 TTTGACATGATTTTGAATTTTCTCTTCAATGTTTACGGCTCTTTTCAAGCAATTAAT 720
Db 50364 TTTGACATGATTTTGAATTTTCTCTTCAATGTTTACGGCTCTTTTCAAGCAATTAAT 50305

QY 721 TTTTCTTACGTGTGATGATTAAGACCAACATACACAGTGTGATGATGATTAAGTT 780
Db 50304 TTTTCTTACGTGTGATGATTAAGACCAACATACACAGTGTGATGATGATTAAGTT 50245

```

```

QY 781 CGGGGGATTAAGCATGATTTCTGTAGAGGATGATCAAAAACTTTGGGCTGTTCTGGAAG 840
Db 50244 CGGGGGATTAAGCATGATTTCTGTAGAGGATGATCAAAAACTTTGGGCTGTTCTGGAAG 50185

QY 841 TGTATTTCTTTTAAAGTGTGCTCTCGGGTTTGGAGGCTGGAATTCGGTTTGGGAAT 900
Db 50184 TGTATTTCTTTTAAAGTGTGCTCTCGGGTTTGGAGGCTGGAATTCGGTTTGGGAAT 50125

QY 901 GGATTTTAAAGAGGGATATGCTTTTACCTTTATCCAAAAGCATGAGATCAGCATAT 960
Db 50124 GGATTTTAAAGAGGGATATGCTTTTACCTTTATCCAAAAGCATGAGATCAGCATAT 50065

QY 961 TGCTCAATGCGTGGCAAGTTGTGATTAACATACAGAAAGCTGGCTTCTCTAGAGA 1020
Db 50064 TGCTCAATGCGTGGCAAGTTGTGATTAACATACAGAAAGCTGGCTTCTCTAGAGA 50005

QY 1021 CTTCGGTATTTCAAACTTTTGTGATCTTCAGAAAAGATCAAAATCTATTTAGTATTAAC 1080
Db 50004 CTTCGGTATTTCAAACTTTTGTGATCTTCAGAAAAGATCAAAATCTATTTAGTATTAAC 49945

QY 1081 TTTAAGCTATACT-AGCGATACAGAGCTCTCTCTTAAATTAACATCATGAGCTGGC 1139
Db 49944 TTTAAGCTATACTTAAGAGATACAGAGCTCTCTCTTAAATTAACATCATGAGCTGGC 49885

QY 1140 GTTA-TTGTGGGATGTGTGCAAAAAGAGGCTAGATTTCTAGCGAAACTTAACGA 1198
Db 49884 GTTAAGTGTGGGATGTGTGCAAAAAGAGGCTAGATTTCTAGCGAAACTTAACGA 49825

QY 1199 AACGCAAAA-TTGTGTCAAGGTGAGAGCAAACTATCGAAGAAAATGCTTATCAGGC 1257
Db 49824 AACGCAAAAATTTTGTGTCAAGGTGAGAGCAAACTATCGAAGAAAATGCTTATCAGGC 49765

QY 1258 GACCATCGGGCTTTTGAAGCTTTGGCAATCATCTTGGCTATGATGATTTGGCTTGA 1317
Db 49764 GACCATCGGGCTTTTGAAGCTTTGGCAATCATCTTGGCTATGATGATTTGGCTTGA 49705

QY 1318 ATGGCAATATGCTTTCAGTGGCGGTATGCGCTTAAATTCATGACCTTTGGCTACCTGTC 1377
Db 49704 ATGGCAATATGCTTTCAGTGGCGGTATGCGCTTAAATTCATGACCTTTGGCTACCTGTC 49645

QY 1378 AGTCTGTTTATAGCAATTTCTTTTGAAGAAATTCAAATGATTTGCAAGCCATTGG 1437
Db 49644 AGTCTGTTTATAGCAATTTCTTTTGAAGAAATTCAAATGATTTGCAAGCCATTGG 49585

QY 1438 TGTCTTAATGACTGTATTTGGGGTATTCATTAACAATATCTTGAATGATTTGATGCTAT 1497
Db 49584 TGTCTTAATGACTGTATTTGGGGTATTCATTAACAATATCTTGAATGATTTGATGCTAT 49525

QY 1498 TCGTGAAGATCGCAAGCCGAACTGTTTACCCTATGCAATGTTTGAATGATGACCT 1557
Db 49524 TCGTGAAGATCGCAAGCCGAACTGTTTACCCTATGCAATGTTTGAATGATGACCT 49465

QY 1558 TCAAAAGACGTTACGCGCAGCGTAATGACACAGCTACAACTATCAAGTTTGTAAAT 1617
Db 49464 TCAAAAGACGTTACGCGCAGCGTAATGACACAGCTACAACTATCAAGTTTGTAAAT 49405

QY 1618 GCTTTGTTTATAGGGGCTCTCTGCTCTTAAATTTTGAATTTATTAAGCAATAGGAT 1677
Db 49404 GCTTTGTTTATAGGGGCTCTCTGCTCTTAAATTTTGAATTTATTAAGCAATAGGAT 49345

QY 1678 TCTTCTAGGAACCTTATGCTCTCTTATATATGACCAACCTGTTGTTGTTATAGGCG 1737
Db 49344 TCTTCTAGGAACCTTATGCTCTCTTATATATGACCAACCTGTTGTTGTTATAGGCG 49285

QY 1738 TAAAGAAAATCGCTCAAAATTAAGTACGTTTAACTTAATCTTAACGTTAGCAATATAAA 1797
Db 49284 TAAAGAAAATCGCTCAAAATTAAGTACGTTTAACTTAATCTTAACGTTAGCAATATAAA 49225

QY 1798 ATCTCTTTGGAGCTTATGATCCCAAGGCCCTGTGTGATTTAAATTTATGACAAATTCAG 1857
Db 49224 ATCTCTTTGGAGCTTATGATCCCAAGGCCCTGTGTGATTTAAATTTATGACAAATTCAG 49165

QY 1858 ATATATGC 1864

```


Query Match	Best Local Similarity	Matches 1831	Conservative	0	Mismatches	1	Indels	3	Gaps	3
61	AAACCGTGATGCGGATGCTGTAGTATGACGGTTATATGACAGAGCCCTATT	120								
DB	10424	AAACCGTGATGCGGATGCTGTAGTATGACGGTTATATGACAGAGCCCTATT	10483							

QY	1199	AAAGCTCCCATTTGAAAAATCATGCAAGTGTCTACGGGAAATTTTACCAACGCTGAAGTAG	180
Db	10484	AAAGCTCCCATTTGAAAAATCATGCAAGTGTCTACGGGAAATTTTACCAACGCTGAAGTAG	10543
QY	181	CAAACTGCCTCAAGATTTTAAATCTGAGCGATGTCCTTTGTTCCGAGGTTCTCACTGA	240
Db	10544	CAAACTGCCTCAAGATTTTAAATCTGAGCGATGTCCTTTGTTCCGAGGTTCTCACTGA	10603
QY	241	AGAGACGATCTCTTCGTGATCTTTGGGAAAAACATGTATCAACAAGGCATTAATCTACATG	300
Db	10604	AGAGACGATCTCTTCGTGATCTTTGGGAAAAACATGTATCAACAAGGCATTAATCTACATG	10663
QY	301	CTGTGAGCTTTGGCAATGCTTATTTGTTTGAATAGCGTATATTAATGATTTTGGAGCGCTCAT	360
Db	10664	CTGTGAGCTTTGGCAATGCTTATTTGTTTGAATAGCGTATATTAATGATTTTGGAGCGCTCAT	10723
QY	361	CGCTTCGGAGCGCTGTCCTTCGTAACTCTTTCCTTATCTGGGACGCTACAGTATTTTGA	420
Db	10724	CGCTTCGGAGCGCTGTCCTTCGTAACTCTTTCCTTATCTGGGACGCTACAGTATTTTGA	10783
QY	421	TGCGGCATCTCACCTTTGTAGAGACTGCGTGGGATGTTTCTTGTATGGGATGGCCGTAGA	480
Db	10784	TGCGGCATCTCACCTTTGTAGAGACTGCGTGGGATGTTTCTTGTATGGGATGGCCGTAGA	10843
QY	481	TGCAAAATGCTTCGTGATTCGAAAAGATCCGAGAGAAATTTTATATGCTCAAGCTTAA	540
Db	10844	TGCAAAATGCTTCGTGATTCGAAAAGATCCGAGAGAAATTTTATATGCTCAAGCTTAA	10903
QY	541	AAAATCTGTAGAAAAAGATATACCAAGGCTTTTGGAGCCATTTTGTATCTTAACCTTGAC	600
Db	10904	AAAATCTGTAGAAAAAGATATACCAAGGCTTTTGGAGCCATTTTGTATCTTAACCTTGAC	10963
QY	601	TACAGTATTGGGCTCAGACATTTCTTTTCTCTAGATACAGGCGCTATTTAAAGGTTTGC	660
Db	10964	TACAGTATTGGGCTCAGACATTTCTTTTCTCTAGATACAGGCGCTATTTAAAGGTTTGC	11023
QY	661	TTTGCATTTGATTTTGAAGATTTTCTCTCAATGTTTACGGGCTCTTTCAAGACTAAAT	720
Db	11024	TTTGCATTTGATTTTGAAGATTTTCTCTCAATGTTTACGGGCTCTTTCAAGACTAAAT	11083
QY	721	TTTCTTCATGCTGTGATGATATAGAACCCCAATACACAGTTGATATGATGATTAAGTT	780
Db	11084	TTTCTTCATGCTGTGATGATATAGAACCCCAATACACAGTTGATATGATGATTAAGTT	11143
QY	781	CGTGGGATTAAGCATGATTTCTTGAGAGATGCAAAAACCTTGGGCTGTTTCTGGAAG	840
Db	11144	CGTGGGATTAAGCATGATTTCTTGAGAGATGCAAAAACCTTGGGCTGTTTCTGGAAG	11203
QY	841	TGTTTTTCTTTTAGGTTGCGGTGCTCTGGGGTTTGGACCTGGGAATTCGTTTTGGGAT	900
Db	11204	TGTTTTTCTTTTAGGTTGCGGTGCTCTGGGGTTTGGACCTGGGAATTCGTTTTGGGAT	11263
QY	901	GGATTTTAAAGGAGGATGTCCTTATCTTTAATCCAAAAGAGCATGGCATCAGCGATGT	960
Db	11264	GGATTTTAAAGGAGGATGTCCTTATCTTTAATCCAAAAGAGCATGGCATCAGCGATGT	11323
QY	961	TGCTCAAAATGCGTGCAAAAGTTTGATATAAATCTACAGGAAGCTGTCCTTCTTACAGA	1020
Db	11324	TGCTCAAAATGCGTGCAAAAGTTTGATATAAATCTACAGGAAGCTGTCCTTCTTACAGA	11383
QY	1021	CTTCCGATTTCAAAATTTTGGATCTTCAGAAAAGATTCAAAATCTATTTTAGATATAAGC	1080
Db	11384	CTTCCGATTTCAAAATTTTGGATCTTCAGAAAAGATTCAAAATCTATTTTAGATATAAGC	11443
QY	1081	TTTAGGCTATACT-AAGCAGATACGAGGCTCTCTCTTAATTTAATCAATCAATGAGCTGGC	1139
Db	11444	TTTAGGCTATACTAATGAGATACGAGGCTCTCTCTTAATTTAATCAATCAATGAGCTGGC	11503
QY	1140	GTTTA-TTGTGGATTTGTGTGCAAAAAGGCGCTAGATTTCTCTACGAAAACCTTAACGA	1198
Db	11504	GTTTACTGTGGGATTTGTTGTGCAAAAAGGCGCTAATTTCTCTACGAAAACCTTAACGA	11563
QY	1199	AAAGCAAAA-TTTTGCTAAAGGTAGACGCAAACTATGAAAGAAATGCGTTATCAGGC	1257


```

/product="glutamyl-tRNA synthetase"
/protein_id="AAF38063.1"
/db_xref="GI:7189121"
/translation="MMENVNRVAVAPPTGDPHVGTAYMALFNEIFAKRFGKGMILRI
EDTRSRHODYEENIFSAIRMGCIQMDGPDVGVGYPYROSERTKIQGVETLLK
TPCAYCFATPOELAEVAVASTLGRGVDPRYRLSPEEVAAREAAQOYITRLKY
PLSGECVFRIDYSGKGRVFPWADVDQVIVKSGFPRYHANYIDHLMITVILGEE
WLSSTPQHLLLEARGMEPPVFLHMFLLNPDGTAKSKRANPTSTFYADSGYKEAP
VNFPLTMGSMGDEEVSLERTIETENPRRIKSGAVDIOQLDMNKHYNHESSP
ECLIKELQGMNLNDEFFLKLPLCGSRITTLAEFINLTSFPFSGLEIYVEELLQAL
SEKAIILYSYVKYLEKTDQMTKCTCYGSKWLAQAFVNHKKAILPILYVALTGKK
OGILPFDSEIIGKPRARALVYAEKLGSGVPRKLAATVDKFMQREDPEAFEDL"
/complement(6199..6387)
/gene="CP0191"
/complement(6199..6387)
/notes="conserved hypothetical protein; identified by
Glimmer2: putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF38064.1"
/db_xref="GI:7189122"
/translation="MKLFYFCNIIPLTSLPSVSCHTPPPRSYTLAQGRPLVNXAHS
LEPFLARSVNTCYNTNL"
/gene="CP0192"
/gene="CP0192"
/6645..6752
/6645..6752
/gene="CP0192"
/notes="hypothetical protein; identified by Glimmer2:
putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF38065.1"
/db_xref="GI:7189123"
/translation="MFSKSKSFNTKTIYLEKTIKIFIFICNVHGVLY"
/6829..7101
/6829..7101
/gene="CP0193"
/6829..7101
/gene="CP0193"
/notes="similar to PID:1783381; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="cysteine rich outer membrane protein 3"
/protein_id="AAF38066.1"
/db_xref="GI:7189124"
/translation="MKKAVYLAMFGGVVSLSSCCRIYVCCFEDPCAPSSCNCEVIR
KERSCGNACGSPSCNPGSTECNSQSPVQKCTSPDKCKQ"
/complement(7165..7281)
/gene="CP0194"
/complement(7165..7281)
/notes="hypothetical protein; identified by Glimmer2:
putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF38067.1"
/db_xref="GI:7189125"
/translation="MSLDIGISYRMFIFPFCRVECFVNSYPOHLGLSELRL"
/7268..8538
/7268..8538
/gene="CP0195"
/7268..8538
/gene="CP0195"
/notes="similar to SP:P23700 PID:550566 PID:1326243
PID:1326245 GB:AE001363; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="60 kDa outer membrane protein"
/protein_id="AAF38068.1"

```

```

/db_xref="GI:7189126"
/translation="MSKLIIRVVYVTLATSMASCPASGIEBAVAESLITKVASAE
KPAVPMATKAVRLVRNKQVEQKSCAFCDKERYPEEERGQDPEAEQSSCTGRLY
SVKANDDCNVBIQOSVPEYATVGSPPYLEIILAIKGDVDVAVITQLECEAFVSSDP
ETTPSPDCKLWKKIDRLGAGDKCKTITWKKPKRECCPTAATVACPELRKYKCGP
AICIKOEPDCACLRCPCYKIEVNTGSAIARNTVNPVPDGVSHAGORVLSFNL
GDMRGDKKVFTEVCEPCORGOITNVAVTYCGGKGCANATVTVNECCOVNINSAD
MSYCKPVEISIVSNPGDVLVHVDITLPSGVTVLEAPGGEICCKKVMRIKEMC
PGETIQFLVYKAVQPGPFTNOAVATSSNCGTCSCAETTHKGLAATMKCULDITN
DPTICGVENIVRICTVNGSADNVNLSILKFSKELQPIASSGPTKGTISGTVFEDA
LPKCSKSVSFVSVTLKQIAPGDARGAIISSDITLSPVSDENTHYV"
/9217..9807
/gene="CP0196"
/9217..9807
/gene="CP0196"
/notes="similar to SP:P18587; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="15 kDa cysteine-rich outer membrane protein"
/protein_id="AAF38069.1"
/db_xref="GI:7189127"
/translation="MSSNMLHPVGGTGTGAAPESVNIIVEBIAASGVYAGLQATSS
PGVMVLIGMAKTKFIOPRESKLFQSRACQITLVGLILVAGLACMFTFHSOLGA
NAFWLITPAIAGLILVLTSLCPDEKACSEKLMVQKAGVLEQDLDDGILNNSKIF
GHVKTENTSRATTPVNDGRTIPVLSVSKIARY"
/complement(9885..11831)

```

gene

Query Match

Best Local Similarity 4.9%; Score 92; DB 1; Length 11938;

Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1773 TAATCTAACGCTGACGATATATAAAATTCCTTGCGACTTATAGTCCAAAGCCCTGT 1832

Db 1 TAATCTAACGCTGACGATATATAAAATTCCTTGCGACTTATAGTCCAAAGCCCTGT 60

Qy 1833 GGTATTAATTTATGCAATTCAGATAATGC 1864

Db 61 GGTATTAATTTATGCAATTCAGATAATGC 92

RESULT 5

AF019977/c 3183 bp mRNA linear VMT 15-NOV-1997

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Gallus gallus.

Gallus gallus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Archosauria; Aves; Neognathae; Galliformes; Phasianidae;

Phasianinae; Gallus.

1 (bases 1 to 3183)

Quik'J', van den Heuvel,M., Henrique,D., Mariño,V., Sheer,D.,

Tabin,C., and Ingham,P.W.

The smoothened gene and Hedgehog Signal Transduction in Drosophila

and Vertebrate Development

Cold Spring Harb. Symp. Quant. Biol. (1997) In press

2 (bases 1 to 3183)

Quik'J', van den Heuvel,M., Henrique,D., Mariño,V., Sheer,D.,

Tabin,C., and Ingham,P.W.

Direct Submision

Submitted (18-AUG-1997) Molecular Embryology Laboratory, Imperial

Cancer Research Fund, 44 Lincoln's Inn Fields, London WC2 3PX,

United Kingdom

On Nov 15, 1997 this sequence version replaced gi:2582358.

COMMENTS

FEATURES

source

1..3183

/organism="Gallus gallus"

/db_xref="taxon:9031"

<1..2387

/note="putative trans membrane protein; similar to


```

repeat_region      /rpt_family="MER66A-int"
complement(7740. .8116)
/rpt_family="THE1C"
repeat_region      complement(8117. .8450)
/rpt_family="MER66A-int"
complement(8533. .8821)
/rpt_family="MER66A-int"
repeat_region      complement(8824. .9097)
/rpt_family="MER66A-int"
complement(9273. .9465)
/rpt_family="MER66B-int"
repeat_region      complement(9482. .9919)
complement(9928. .10257)
/rpt_family="MER66B-int"
repeat_region      complement(9928. .10257)
/rpt_family="MER66B"
repeat_region      /rpt_family="MER66B"
10385. .10634
/rpt_family="HAL1b"
repeat_region      /rpt_family="L1M4"
11492. .12021
/rpt_family="L1M4"
repeat_region      /rpt_family="AT_rich"
12057. .12098
complement(12123. .12199)
/rpt_family="MER5B"
repeat_region      12247. .13722
/rpt_family="L1M4A"
13723. .13927
/rpt_family="L1M4A"
repeat_region      complement(13928. .14234)
/rpt_family="A1y"
14235. .14575
/rpt_family="L1M4A"
repeat_region      /rpt_family="L1M4A"
14608. .14754
/rpt_family="L1M4A"
repeat_region      14757. .14825
/rpt_family="L1M4A"
repeat_region      14832. .14879
/rpt_family="AT_rich"
15041. .15400
/rpt_family="L1M4A"
repeat_region      15674. .15748
/rpt_family="MIR"
17417. .17437
/rpt_family="AT_rich"
repeat_region      complement(17759. .18249)
/rpt_family="L1M4A"
repeat_region      complement(18656. .18809)
/rpt_family="FRAM"
18878. .18905
/rpt_family="AT_rich"
repeat_region      complement(19369. .19479)
/rpt_family="L1M4A"
repeat_region      complement(19841. .20171)
/rpt_family="MER21B"
20172. .20498
/rpt_family="AT_rich"
repeat_region      complement(20499. .20538)
/rpt_family="MER21B"
repeat_region      complement(20539. .21022)
/rpt_family="MER66B"
repeat_region      complement(21023. .21485)
/rpt_family="MER21B"
repeat_region      22732. .22820
complement(22990. .23052)
/rpt_family="L1M4"
24042. .24067
/rpt_family="AT_rich"
repeat_region      24261. .24921
/rpt_family="MER21B"
repeat_region      complement(26131. .26365)
/rpt_family="MIR"
26415. .26475
/rpt_family="MIR"

```

```

repeat_region      26691. .26942
/rpt_family="L1M4"
repeat_region      complement(27050. .27100)
/rpt_family="L1M4"
repeat_region      27183. .27344
/rpt_family="L1M4"
repeat_region      27472. .27492
/rpt_family="AT_rich"
repeat_region      27600. .27802
/rpt_family="L1M4"
repeat_region      27813. .28068
/rpt_family="L1M4"
repeat_region      28507. .28949
/rpt_family="MSTB2"

```

```

Query Match      1.2% Score 22; DB 9; Length 151032;
Best Local Similarity 100.0%; Pred. No. 2.7;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1395 ATTCTTTTGAAGAAATTCA 1416
Db 19872 ATTCTTTTGAAGAAATTCA 19851

```

```

RESULT 7
AC124363/c 159198 bp DNA linear HTG 20-JUN-2002
LOCUS Mus musculus chromosome UNK clone RP24-531A9, WORKING DRAFT
DEFINITION SEQUENCE, 1 unordered piece.
AC124363
AC124363.1 GI:21426484
VERSION HTG; HTGS PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
KEYWORDS house mouse.
SOURCE Mus musculus.
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

```

```

REFERENCE 1 (bases 1 to 159198)
AUTHORS McPherson,J.D. and Waterston,R.H.
TITLE The sequence of Mus musculus clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 159198)
AUTHORS McPherson,J.D. and Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (14-JUN-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
REFERENCE 3 (bases 1 to 159198)
AUTHORS McPherson,J.D. and Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (20-JUN-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
COMMENT
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@wustl.edu
----- Project Information -----
Center project name: M_BB0531A09
----- Summary Statistics -----
Sequencing vector: M13; 0%
Sequencing vector: plasmid; 100%
Chemistry: Dye-terminator Big Dye; 0% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 158388 bases at least Q40
Consensus quality: 158639 bases at least Q30
Consensus quality: 158759 bases at least Q20
Insert size: 13800; agarose-fp
Insert size: 159198; sum-of-ctnigs
Quality coverage: 15.36 in Q20 bases; agarose-fp
Quality coverage: 13.93 in Q20 bases; sum-of-ctnigs

```


* NOTE: This is a 'working draft' sequence. It currently
 * consists of 1 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 159198: contig of 159198 bp in length.
 Location/Qualifiers
 1. 159198
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /chromosome="XNK"
 /clones="RP24-531A9"
 misc_feature
 1. 159198
 /note="assembly name:Contig8"
 BASE COUNT 44143 a 34841 c 35294 g 44919 t 1 others
 ORIGIN

Query Match 1.2%; Score 22; DB 2; Length 159198;
 Best Local Similarity 100.0%; Pred. No. 2.7;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 831 TTTCTGGAAGTCTTTTCTTTT 852
 Db 39433 TTTCTGGAAGTCTTTTCTTTT 39412

RESULT 8
 AC036125 172579 bp DNA linear PRI 24-AUG-2002
 LOCUS
 DEFINITION Homo sapiens chromosome 18, clone RP11-482F20, complete sequence.
 AC036125
 VERSION
 AC036125.7 GI:17223368
 KEYWORDS
 HTG
 SOURCE
 human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE
 1 (bases 1 to 172579)
 Birren, B., Linton, L., Nusbaum, C. and Lander, E.
 HOMO SAPIENS CHROMOSOME 18, CLONE RP11-482F20
 JOURNAL
 Unpublished
 REFERENCE
 2 (bases 1 to 172579)
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
 Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F.,
 Boguslavsky, L., Boukhgalter, B., Brown, A., Burkett, G.,
 Campopiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S.,
 Collymore, A., Cooke, P., DeArellano, K., Dewar, K., Diaz, J.S.,
 Dodge, S., Domino, M., Doyle, M., Ferreira, P., FitzHugh, W., Gage, D.,
 Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,
 Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,
 Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
 Klein, J., LaRocque, K., Lamazares, R., Landers, T., Lehoczy, J.,
 Levine, R., Lieu, C., Liu, C., Locke, K., Macdonald, P., Marquis, N.,
 McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R.,
 Meldrim, J., Meneus, L., Mihova, T., Miranda, C., Mlenga, V., Morrow, J.,
 Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
 O'Neill, D., Oliver, J., Pless, J., Peterson, K., Peterson, K., Rotherman, D.,
 Pless, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
 Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
 Stange-Thomann, N., Subramanian, A., Talamas, J.,
 Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J.,
 Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,
 Young, G., Zainoun, J., Zimmer, A. and Zody, M.

Direct Submission
 Submitted (07-APR-2000) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 REFERENCE
 3 (bases 1 to 172579)
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N.,
 Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Boukhgalter, B.,
 Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B.,

TITLE JOURNAL REFERENCE AUTHORS

Direct Submission
 Submitted (01-DEC-2001) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 REFERENCE
 4 (bases 1 to 172579)

Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
 Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Boukhgalter, B.,
 Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B.,
 Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, A.,
 Cooke, P., DeArellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S.,
 Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S.,
 Ginde, S., Goyette, M., Graham, L., Grand-Pierre, N.,
 Jones, C., Kann, L., Karatas, A., Kells, C., LaRocque, K.,
 Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G.,
 Maclean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C.,
 McCarthy, M., McEwan, P., McKernan, K., McPheeters, R., Meldrim, J.,
 Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C.,
 Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D.,
 Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,
 Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,
 Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupack, R.,
 Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
 Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
 Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,
 Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G.,
 Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission
 Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Dec 1, 2001 this sequence version replaced gi:16949121.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 Center project name: L7322
 Center clone name: 482_F_20

FEATURES source

1. 172579
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="4"
 /map="4"
 /clone="RP11-482F20"
 /clone_lib="RP11-11 Human Male BAC"
 1205..1667
 /rpt_family="MER21B"
 repeat_region
 1668..2151
 /rpt_family="MER66B"

TITLE JOURNAL REFERENCE AUTHORS

Direct Submission
 Submitted (07-APR-2000) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 REFERENCE
 3 (bases 1 to 172579)
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
 Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Boukhgalter, B.,
 Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B.,

Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, A.,
 Cooke, P., DeArellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S.,
 Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S.,
 Ginde, S., Goyette, M., Graham, L., Hulme, W., Iliev, I., Johnson, R.,
 Jones, C., Kamat, A., Karatas, A., Kells, C., LaRocque, K.,
 Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G.,
 Maclean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C.,
 McCarthy, M., McEwan, P., McKernan, K., McPheeters, R., Meldrim, J.,
 Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C.,
 Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D.,
 Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,
 Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,
 Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupack, R.,
 Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
 Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
 Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,
 Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G.,
 Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

```

repeat_region      2152..2191
/rpt_family="MER21B"
complement(2192..2517)
repeat_region      /rpt_family="AluSx"
2518..2848
/rpt_family="MER21B"
repeat_region      3210..3320
/rpt_family="L1PB3"
3784..3812
/rpt_family="(A)n"
repeat_region      3881..4034
/rpt_family="FRAM"
4441..4931
/rpt_family="LOR1a"
5253..5273
/rpt_family="AT_rich"
repeat_region      complement(6942..7016)
/rpt_family="MIR"
repeat_region      complement(7290..7649)
/rpt_family="L1ME4A"
7811..7858
/rpt_family="AT_rich"
complement(7865..7933)
repeat_region      /rpt_family="L1MA4"
/rpt_family="L1R27"
repeat_region      complement(7936..8082)
/rpt_family="L1R27"
repeat_region      complement(8115..8455)
/rpt_family="L1R27"
8456..8762
/rpt_family="AluY"
complement(8763..8967)
repeat_region      /rpt_family="L1R27"
complement(8968..10443)
/rpt_family="L1MA4"
10491..10567
/rpt_family="MER5B"
10592..10633
/rpt_family="AT_rich"
complement(10669..11198)
repeat_region      /rpt_family="L1M4"
complement(12056..12305)
/rpt_family="HAL1b"
12433..12762
/rpt_family="MER6B"
12771..13208
/rpt_family="MER6B-int"
13325..13417
/rpt_family="MER6A-int"
13593..13866
/rpt_family="MER6A-int"
13869..14157
/rpt_family="MER6A-int"
14240..14573
/rpt_family="MER6A-int"
14574..14950
/rpt_family="MER6A-int"
14951..15490
/rpt_family="TBE1C"
15491..15576
/rpt_family="MER6A-int"
15577..15807
/rpt_family="MER6A-int"
16025..16375
/rpt_family="MER6A-int"
16413..16824
/rpt_family="MER6A"
16966..16994
/rpt_family="AT_rich"
complement(17062..17359)
/rpt_family="AluSg"
17670..17692
/rpt_family="AT_rich"
17719..17753

```

```

repeat_region      /rpt_family="(CA)n"
complement(18579..18870)
/rpt_family="AluSx"
repeat_region      20083..20107
/rpt_family="(CA)n"
repeat_region      21436..21613
/rpt_family="L1MD2"
repeat_region      22014..22159
/rpt_family="(TA)n"
repeat_region      complement(22160..22441)
/rpt_family="AluSc"
22776..23232
repeat_region

Query Match      1.2%: Score 22; DB 9; Length 172579;
Best Local Similarity 100.0%: Pred. No. 2.7;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1395 ATTTCTTTTGAAGAAATTC A 1416
Db 2817 ATTTCTTTTGAAGAAATTC A 2838

RESULT 9
AC121084
LOCUS
DEFINITION
Mus musculus clone RP23-191E14, WORKING DRAFT SEQUENCE, 21 ordered
pieces.
AC121084
VERSION
AC121084.3 GI:20986564
KEYWORDS
HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE
Mus musculus.
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 209636)
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Mus musculus, clone RP23-191E14
2 (bases 1 to 209636)
REFERENCE
Birken, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L.,
Boukngalter, B., Brown, A., Camarata, J., Campiano, A., Chang, J.,
Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A.,
Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S.,
Faro, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S.,
Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
Kamat, A., Karatas, A., Kells, C., Larocque, K., Lamazares, R.,
Landers, T., Lehoczeky, J., Levine, R., Lindblad-Toh, K., Liu, G.,
Maclean, C., Macdonald, P., Major, J., Margulis, N., Matthews, C.,
McCarthy, M., McEwan, P., McKernan, K., Meldrum, J., Menus, L.,
Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R.,
Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D.,
Oliyer, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,
Raymond, C., Retta, R., Rieback, M., Riley, R., Riese, C., Rogov, P.,
Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupack, R.,
Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Straus, N., Subramanian, A., Talams, J., Teste, S., Theodore, J.,
Topham, K., Travers, M., Travis, N., Triggillo, V., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G.,
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
Direct Submission
Submitted (15-MAY-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 209636)
REFERENCE
Birken, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L.,
Boukngalter, B., Brown, A., Camarata, J., Campiano, A., Chang, J.,
Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A.,
Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S.,
Faro, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S.,
Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
Hagos, B., Horton, L., Hulme, W., Iliev, I.,

```

Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., LaRocque, K., Lanazares, R., Landers, T., Lehoczy, J., Levine, R., Lindblad-Toh, K., Liu, G., MacLean, C., MacDonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrum, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupack, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zemdek, L., Zimmer, A. and Zody, W.

Direct Submission
Submitted (20-MAY-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On May 20, 2002 this sequence version replaced gi:20901891.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

TITLE
JOURNAL
COMMENT

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIGR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submission@genome.wi.mit.edu
----- Project Information
Center project name: L21012
Center clone name: 191.E.14
----- Summary Statistics

Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 202259 bases at least Q40
Consensus quality: 205316 bases at least Q30
Consensus quality: 206630 bases at least Q20
Insert size: 207636; sum-of-contigs
Quality coverage: 7.6 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 21 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* been provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

* 1 12959: contig of 12959 bp in length
* 12960 13059: gap of 100 bp
* 13060 13748: contig of 689 bp in length
* 13749 13848: gap of 100 bp
* 13849 14975: contig of 1127 bp in length
* 14976 15075: gap of 100 bp
* 15076 16142: contig of 1067 bp in length
* 16143 16242: gap of 100 bp
* 16243 17810: contig of 1568 bp in length
* 17811 17910: gap of 100 bp
* 17911 18957: contig of 1047 bp in length
* 18958 19057: gap of 100 bp
* 19058 20469: contig of 1412 bp in length
* 20470 20569: gap of 100 bp
* 20570 22600: contig of 2031 bp in length
* 22601 22700: gap of 100 bp
* 22701 24776: contig of 2076 bp in length
* 24777 24876: gap of 100 bp
* 24877 29447: contig of 4571 bp in length
* 29448 29547: gap of 100 bp
* 29548 38261: contig of 8714 bp in length
* 38262 38361: gap of 100 bp
* 38362 46977: contig of 8616 bp in length
* 46978 47077: gap of 100 bp
* 47078 57145: contig of 10068 bp in length

* 57146 57245: gap of 100 bp
* 57246 68889: contig of 11644 bp in length
* 68890 68989: gap of 100 bp
* 68990 82201: contig of 13212 bp in length
* 82202 82301: gap of 100 bp
* 82302 97077: contig of 14776 bp in length
* 97078 97177: gap of 100 bp
* 97178 109416: contig of 12239 bp in length
* 109417 109516: gap of 100 bp
* 109517 124991: contig of 15475 bp in length
* 124992 125091: gap of 100 bp
* 125092 151033: contig of 25942 bp in length
* 151034 151333: gap of 100 bp
* 151334 185034: contig of 33901 bp in length
* 185035 185134: gap of 100 bp
* 185135 209636: contig of 24502 bp in length.

FEATURES

source
1. .209636
Location/Qualifiers
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="RP23-191E14"
/clone_lib="RPC1-23 Female Mouse BAC"

misc_feature

1. .12959
/note="assembly_fragment"
clone_end:SP6
vector_side:left

misc_feature

13060. .13748

/note="assembly_fragment"

misc_feature

13849. .14975

/note="assembly_fragment"

misc_feature

15076. .16142

/note="assembly_fragment"

misc_feature

16243. .17810

/note="assembly_fragment"

misc_feature

17911. .18957

/note="assembly_fragment"

misc_feature

19058. .20469

/note="assembly_fragment"

misc_feature

20570. .22600

/note="assembly_fragment"

misc_feature

22701. .24776

/note="assembly_fragment"

misc_feature

24877. .29447

/note="assembly_fragment"

misc_feature

29548. .38261

/note="assembly_fragment"

misc_feature

38362. .46977

/note="assembly_fragment"

misc_feature

47078. .57145

/note="assembly_fragment"

misc_feature

57246. .68889

/note="assembly_fragment"

misc_feature

68990. .82201

/note="assembly_fragment"

misc_feature

82302. .97077

/note="assembly_fragment"

misc_feature

97178. .109416

/note="assembly_fragment"

misc_feature

109517. .124991

/note="assembly_fragment"

misc_feature

125092. .151033

/note="assembly_fragment"

misc_feature

151134. .185034

/note="assembly_fragment"

misc_feature

185135. .209636

/note="assembly_fragment"

clone_end:T7
vector_side:right

BASE COUNT 58935 a 45360 c 45624 g 57716 t 2001 others
ORIGIN

Query Match 1.2%; Score 22; DB 2; Length 209636;
Best Local Similarity 100.0%; Pred. No. 2.7;

Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 831 TTTCTGGAAGTGTTCCTTT 852
Db 88632 TTTCTGGAAGTGTTCCTTT 88653

RESULT 10
AC121990 247074 bp DNA linear HTG 21-MAY-2002
LOCUS Mus musculus chromosome UNK clone RP24-298L8, WORKING DRAFT
DEFINITION SEQUENCE, 15 unordered pieces.
AC121990
AC121990.1 GI:21040114
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
KEYWORDS Mus musculus.
SOURCE Mus musculus.
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 247074)
AUTHORS McPherson,J.D. and Waterston,R.H.
JOURNAL The sequence of Mus musculus clone
TITLE Unpublished
REFERENCE 2 (bases 1 to 247074)
AUTHORS McPherson,J.D. and Waterston,R.H.
JOURNAL Direct Submission
TITLE Submitted (21-MAY-2002) Genome Sequencing Center, 4444 Forest Park
JOURNAL Parkway, St. Louis, MO 63108, USA

COMMENT
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@wustl.wustl.edu
----- Project Information -----
Center project name: M_BB0298L08
----- Summary Statistics -----
Sequencing vector: M13; 0%
Sequencing vector: plasmid; 100%
Chemistry: Dye-terminator Big Dye; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.99019
Consensus quality: 240193 bases at least Q40
Consensus quality: 241582 bases at least Q30
Consensus quality: 242647 bases at least Q20
Insert size: 18300; agarose-fp
Insert size: 245674; sum-of-contigs
Quality coverage: 9.96 in Q20 bases; agarose-fp
Quality coverage: 6.54 in Q20 bases; sum-of-contigs
----- NOTE: This is a 'working draft' sequence. It currently consists of 15 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved. -----
1 1289: contig of 1289 bp in length
* 1280 1389: gap of unknown length
* 1390 3596: contig of 2207 bp in length
* 3597 3696: gap of unknown length
* 3697 5665: contig of 1969 bp in length
* 5666 5765: gap of unknown length
* 5766 9154: contig of 3389 bp in length
* 9155 9254: gap of unknown length
* 9255 14184: contig of 4930 bp in length
* 14185 14284: gap of unknown length
* 14285 19239: contig of 4955 bp in length
* 19240 19339: gap of unknown length
* 19340 25026: contig of 5687 bp in length
* 25027 25126: gap of unknown length

FEATURES
source
1. 247074
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="UNK"
/clone="RP24-298L8"
1. 1289
/note="assembly_name:Contig15"
misc_feature 1390..3596
/note="assembly_name:Contig16"
misc_feature 3697..5665
/note="assembly_name:Contig17"
misc_feature 5766..9154
/note="assembly_name:Contig18"
misc_feature 9255..14184
/note="assembly_name:Contig19"
misc_feature 14285..19239
/note="assembly_name:Contig20"
misc_feature 19340..25026
/note="assembly_name:Contig21"
misc_feature 25127..35457
/note="assembly_name:Contig22"
misc_feature 35558..45385
/note="assembly_name:Contig23"
misc_feature 45486..60818
/note="assembly_name:Contig24"
misc_feature 60919..79229
/note="assembly_name:Contig25"
misc_feature 79330..104699
/note="assembly_name:Contig26"
misc_feature 104800..133578
/note="assembly_name:Contig27"
misc_feature 133679..171553
/note="assembly_name:Contig28"
misc_feature 171654..247074
/note="assembly_name:Contig29"

BASE COUNT 66398 a 54384 c 55492 g 63387 t 1413 others
ORIGIN

Query Match 1.2% Score 22; DB 2; Length 247074;
Best Local Similarity 100.0%; Pred. No. 2.7;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 979 AGTTGTCATTAACCTACAGGAA 1000
Db 202271 AGTTGTCATTAACCTACAGGAA 202292

RESULT 11
AC120685/c 293272 bp DNA linear HTG 24-AUG-2002
LOCUS Rattus norvegicus clone CH230-13J13, *** SEQUENCING IN PROGRESS
DEFINITION *** 118 unordered pieces.
AC120685
AC120685.3 GI:22164888
HTG; HTGS_PHASE1.
KEYWORDS Norway rat.
SOURCE Rattus norvegicus
ORGANISM Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE AUTHORS

1 (bases 1 to 293272)
Muzny, D., Marie, E., Metzger, M., Lee, S., Abramson, S., Adams, C., Alder, J., Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Bielawski, K., Blair, C., Blanchburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Cesar, H., Chen, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpach, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, D., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorensuhewa, L., Loulseghe, H., Lozada, R.J., Lu, X., Ma, J., Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhinney, S., McLeod, M., McNeill, T., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munitadasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwaokemele, O., Okwuonu, G., Olarunpungu, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poindexter, A., Popovic, D., Pramus, E., Pu, L.-L., Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sleson, I., Sitter, C.D., Smajs, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, P., Williams, G., Willson, R., Wleczyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

TITLE REFERENCE

2 (bases 1 to 293272)
Worley, K.C.

TITLE AUTHORS

Direct Submission
Submitted (09-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

TITLE REFERENCE

3 (bases 1 to 293272)
Rat Genome Sequencing Consortium.

TITLE AUTHORS

Direct Submission
Submitted (24-AUG-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

TITLE COMMENT

On Aug 9, 2002 this sequence version replaced gi:21902829.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information

Center project name: GXNF
Center clone name: CH230-13J13
----- Summary Statistics

Sequencing vector: Plasmid;

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 180898 bases at least Q40

Consensus quality: 191186 bases at least Q30

Consensus quality: 198854 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).

* NOTE: This is a 'working draft' sequence. It currently

* consists of 118 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

* 1
* 1139: contig of 1139 bp in length
* 1140: gap of unknown length
* 1239: gap of unknown length
* 1240: contig of 1182 bp in length
* 2421: gap of unknown length
* 2422: gap of unknown length
* 2521: gap of unknown length
* 2522: gap of unknown length
* 3577: contig of 1056 bp in length
* 3578: gap of unknown length
* 3579: gap of unknown length
* 5345: contig of 1668 bp in length
* 5346: gap of unknown length
* 5446: gap of unknown length
* 6897: contig of 1452 bp in length
* 6898: gap of unknown length
* 6997: gap of unknown length
* 8641: contig of 1644 bp in length
* 8642: gap of unknown length
* 8741: gap of unknown length
* 10061: contig of 1320 bp in length
* 10161: gap of unknown length
* 11497: contig of 1316 bp in length
* 11498: gap of unknown length
* 11597: gap of unknown length
* 12721: contig of 1124 bp in length
* 12722: gap of unknown length
* 12821: gap of unknown length
* 14450: contig of 1629 bp in length
* 14451: gap of unknown length
* 15839: contig of 1289 bp in length
* 15840: gap of unknown length
* 15939: gap of unknown length
* 17030: contig of 1091 bp in length
* 17031: gap of unknown length
* 17131: gap of unknown length
* 18367: contig of 1237 bp in length
* 18368: gap of unknown length
* 18468: gap of unknown length
* 19767: contig of 1300 bp in length
* 19768: gap of unknown length
* 19868: gap of unknown length
* 21091: contig of 1224 bp in length
* 21092: gap of unknown length
* 21191: gap of unknown length
* 22562: contig of 1371 bp in length
* 22563: gap of unknown length
* 23663: contig of 1005 bp in length
* 23668: gap of unknown length
* 23767: gap of unknown length
* 25100: contig of 1333 bp in length
* 25101: gap of unknown length
* 26729: contig of 1529 bp in length
* 26730: gap of unknown length
* 26731: gap of unknown length
* 26829: gap of unknown length
* 28078: contig of 1249 bp in length
* 28079: gap of unknown length
* 28178: gap of unknown length
* 29528: contig of 1350 bp in length
* 29529: gap of unknown length
* 29628: gap of unknown length
* 29629: gap of unknown length
* 31078: contig of 1450 bp in length
* 31079: gap of unknown length
* 32315: contig of 1137 bp in length
* 32415: gap of unknown length
* 33847: contig of 1432 bp in length
* 33848: gap of unknown length
* 33948: contig of 1682 bp in length
* 35629: gap of unknown length
* 35630: gap of unknown length
* 35729: gap of unknown length
* 36762: contig of 1033 bp in length
* 36763: gap of unknown length
* 38277: contig of 1415 bp in length

```

* 38278 38377: gap of unknown length
* 38378 39596: contig of 1219 bp in length
* 39596 39697: gap of unknown length
* 39697 40760: contig of 1064 bp in length
* 40761 40860: gap of unknown length
* 40861 42420: contig of 1560 bp in length
* 42421 42520: gap of unknown length
* 42521 43895: contig of 1375 bp in length
* 43896 43995: gap of unknown length
* 43996 45404: contig of 1409 bp in length
* 45405 47096: contig of 1592 bp in length
* 47097 47197: gap of unknown length
* 47197 48518: contig of 1322 bp in length
* 48519 48619: gap of unknown length
* 48619 49788: contig of 1170 bp in length
* 49789 49888: gap of unknown length
* 49889 51110: contig of 1222 bp in length
* 51111 51210: gap of unknown length
* 51211 53062: contig of 1852 bp in length
* 53063 53162: gap of unknown length
* 53163 54912: contig of 1750 bp in length
* 54913 55012: gap of unknown length
* 55013 56564: contig of 1552 bp in length
* 56565 56664: gap of unknown length
* 56665 57895: contig of 1231 bp in length
* 57896 57995: gap of unknown length
* 57996 59279: contig of 1284 bp in length
* 59280 61376: gap of unknown length
* 61377 61476: gap of unknown length
* 61477 63550: contig of 2074 bp in length
* 63551 63650: gap of unknown length
* 63651 65479: contig of 1829 bp in length
* 65480 65579: gap of unknown length
* 65580 67124: contig of 1545 bp in length
* 67125 67224: gap of unknown length
* 67225 68818: contig of 1594 bp in length
* 68819 68918: gap of unknown length
* 68919 70287: contig of 1369 bp in length
* 70288 70387: gap of unknown length
* 70388 71971: contig of 1584 bp in length
* 71972 72071: gap of unknown length
* 72072 73921: contig of 1850 bp in length
* 73922 74021: gap of unknown length
* 74022 75243: contig of 1222 bp in length

```

Query Match 1.2%; Score 22; DB 2; Length 293272;
 Best Local Similarity 100.0%; Pred. No. 2.7;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 979 AGTTGTCATAACTACAGGAA 1000
 Db 265159 AGTTGTCATAACTACAGGAA 265138

RESULT 12
 AF026147
 LOCUS
 DEFINITION Bacillus subtilis yojA (yojA), yojB (yojB), yojC (yojC), yojD (yojD), yojE (yojE), yojF (yojF), yojG (yojG), yojH (yojH), yojI (yojI), yojJ (yojJ), yojK (yojK), yojL (yojL), yojM (yojM), yojN (yojN), and yojO (yojO) genes, complete cds; and Ocha (ocha) gene, partial cds.
 AF026147
 AF026147.1 GI:3169316

ACCESSION
 VERSION
 KEYWORDS
 ORGANISM
 SOURCE
 REFERENCE
 AUTHORS
 TITLE
 DNA sequences of a 15.4 kb fragment covering the 181 degree region

JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 FEATURES
 source
 of the Bacillus subtilis genome
 Unpublished
 2 (bases 1 to 15389)
 Park, S.-H., Shin, B.-S. and Choi, S.-K.
 Direct Submission
 Submitted (22-SEP-1997) Applied Microbiology Research Division,
 Korea Research Institute of Bioscience and Biotechnology, P.O.Box
 115, Yusong, Taejeon 305-600, Korea
 Location/Qualifiers
 1..15389
 /organism="Bacillus subtilis"
 /strain="168"
 /db_xref="taxon:1423"
 /map="181 degrees"
 /note="sequence upstream of the ocha gene"
 7..1341
 /gene="yojA"
 7..1341
 /note="similar to Bacillus subtilis gluconate permease
 encoded by the sequence presented in GenBank Accession
 Number J02584"
 /codon_start=1
 /transl_table=1
 /product="yojA"
 /protein_id="AAC17849.1"
 /db_xref="GI:3169317"
 /translation="MPILVANGVILFLIKVKLNTFVSLIVSPLVATGLNDIN
 KVLISLETIGQGLHIALVFGMLAGLVSVDAGGYRIATLLIDKGRKRIQAAV
 VASFLIGLIFPEVGLVLIPIVIAIAKELKMPFLYLGIPMAALNVHGFLEPHAP
 TALSVAYNGIQLVLFGLIIIVPTVLAGLFPFNFKMRPGAVQKGNLSGLSPRK
 EROLDETPEGFAISATLSPVIFEMATVIFSLSEHSGKDIIEFIGPTAMLSL
 ILAATMGVARKISMOEIGRSISISQIAAMLLIIGGAGKOVYLDGGVDYADF
 FQIQNSPLPVATVIAVIRLGLSATVAALLTAAAMAPLMAGSVNPLAVLAGAG
 SYIACHVNDAGVMWMEFFGLSKETFTWTLITVLTGCVLAGLVV"
 1426..1662
 /gene="yojB"
 1426..1662
 /gene="yojB"
 /function="unknown"
 /codon_start=1
 /transl_table=1
 /product="yojB"
 /protein_id="AAC17850.1"
 /db_xref="GI:3169318"
 /translation="MPHHSYLRGIPGPGYPARSPFLGALVGGLLGFLGSLFN
 YSRPVAYPPGPGYGGGPGYFGAGVPYGGPGY"
 1770..2171
 /gene="yojC"
 complement(1770..2171)
 /function="unknown"
 /codon_start=1
 /transl_table=1
 /product="yojC"
 /protein_id="AAC17851.1"
 /db_xref="GI:3169319"
 /translation="MPENKRYPPHNOSEKELAKKADPKVEVTVNUNVESVGSNYA
 AQPFPDLPDQKHPAKDPKVPKPDIDFTADHVFAVKVPISEKMLQVARIKATSH
 EIMLEMLPRAESEKSPSLFSQTORNNKSL"
 2270..3061
 /gene="yojD"
 2270..3061
 /gene="yojD"
 /function="unknown"
 /note="similar to Escherichia coli RaxD protein:
 Swiss-Prot Accession Number P27844"
 /codon_start=1
 /transl_table=1
 /product="yojD"
 /protein_id="AAC17852.1"
 /db_xref="GI:3169320"
 /translation="MQHAETKGVITYTAVSFIWMGLFPLYWKLLEQUPALDIAMRII
 MSFVPMCVILFPLRWKIGIQMQLSRKNGGILSLFLASILISINFFVIYIAVNHGFL

```

LEASLGVYINPLVSLGILFLKEKLNELQVLAVSIAAGVTIISAFQVGSIPYVALLL
AFSPCLVLSKRTSLPSAIGLTLETMMIMIALGYLLFSPHQVAPAGAESGCTWLLLF
LAGUPTALPLLLFAGAKRLPLYQVILQYIAPTITLLIGLFCUP"
3243..3521
/gene="yojE"
CDS
3243..3521
/gene="yojE"
/function="unknown"
/codon_start=1
/transl_table=11
/product="YoJE"
/protein_id="AAC17853.1"
/db_xref="GI:3169321"
/translacion="MDNDKEIVKAYASLWNNRSLAHDDAEVAEAILDLELDKXTHPR
LRKPMLEKYPFAIQRIYVNSQLEPAVKYQLVKLHTEABEILKEERGEQS"
3518..3868
/gene="yojF"
CDS
3518..3868
/gene="yojF"
/function="unknown"
/codon_start=1
/transl_table=11
/product="YoJF"
/protein_id="AAC17854.1"
/db_xref="GI:3169322"
/translacion="MKAILKEDVQASLERYADRPVYIHLETTTGSYSAHLNEKNMTVV
AYIRNAKTYHQAKIKGNGPIRVGLKTEEGWIYAEGLUXTYVDBENRLLMAGHLPGGK
LAISLQISEKPFIV"
4122..4550
/gene="yojG"
CDS
4122..4550
/gene="yojG"
/function="unknown"
/codon_start=1
/transl_table=11
/product="YoJG"
/protein_id="AAC17855.1"
/db_xref="GI:3169323"
/translacion="MLGURKTLFEDEYVLIADIMEEIIDVVKPSLIVTFYPGHGVHP
DHDACQTEGMRLREBEKLNPNVMTATPEEIIYTYQWND"
complement(4576..5424)
/gene="yojH"
CDS
complement(4576..5424)
/gene="yojH"
/function="unknown"
/note="similar to Bacillus subtilis YghA protein encoded
by the sequence presented in GenBank Accession Number
D84432 or Swiss-Prot Accession Number: P54504"
/codon_start=1
/transl_table=11
/product="YoJH"
/protein_id="AAC17856.1"
/db_xref="GI:3169324"
/translacion="WAKNKLFEYLSQHAETISSTWYETIETDPNSIYASTDPVVIH
NLKSNLAFNYKINRIFIDDDVILPKLWAFVETQDEHLKPIHIYIREFVRVD
LYVSVEFVHLNQTVKSEAEADYHALIKAFDLVHIIFEEMKNTSLQLOQKDM
ITELSAFVILFHSVGLPIGIDTVRAKLIMENTLHQCAKKVOTLYIDLSGAVI
DTMVAHQLSLIEALRLIGVSTLSGIRPEIAQTAQVQLGLSFEGLSRSTLASIASD
LKLKKV"
complement(5440..6798)
/gene="yoji"
CDS
complement(5440..6798)
/gene="yoji"
/function="unknown"
/note="similar to Escherichia coli hypothetical YdhE
protein and to Haemophilus influenzae hypothetical HI1612
protein encoded by the sequences presented in GenBank
Accession Numbers U68703 and U32834 respectively"
/codon_start=1
/transl_table=11
/product="Yoji"
/protein_id="AAC17857.1"

```

```

/db_xref="GI:3169325"
/translacion="MKETQNAEKIKQLHLHILPFIPTQAGLSLITFLDTVMSKVSVP
ADLAGVAIGSLMTPTVTYGLAGLMAVTPVAQLLGAEKQKIPTVTLQAVYVAALIS
IADVIGYAAVDLILGLRLNDIHHQIAKHLFGLSLGIFPLFTVTLVRSFIDSLGKT
RVMTMITLSSLPINFLVNVVIFGKFGMPALGVGAGLALTVMCICISFFIHKH
APSEYGIFLTMYKFSWKACKNLLKIGLPIGPVAFVETTSIFAATVLLMSHPHTVTIAS
HOAMNPFASLLYMLPLPSVVALTIVGFEAGARFKDARSYSLIGIMMAIGFSLFTAA
CILLFREQIAGMTSDDDVLRLTQHFLIYAFQLSDAVAAPIQGALRGYKDVNTYLLA
AAFVSYWVIGLVPVGMVGTFTSLGAFGYWIGTLAAGAVGLFPRLAKLQKRYSQKQ
HI"
gene
7072..7668
/gene="yojJ"
CDS
7072..7668
/gene="yojJ"
/function="unknown"
/codon_start=1
/transl_table=11
/product="YoJJ"
/protein_id="AAC17858.1"
/db_xref="GI:3169326"
/translacion="MRYEQISENAFKGKIQVYLEQILGDASLILKTLHEKDQCLLCEL
DDLGHVPQDMQGIASSFYQSYIEEFTPAFIELAKAIKALSEHKHGALIVIERADPVE
RFIQKGTSLHABISSLSLIESIFRPGNPLHDLAVRENKLVSASNVLPITTKVEVDIHL
GTRHLRLVCAIRMHWCWSEFQKQRCRLPKTYIIR"
8274..9137
/gene="yojK"
CDS
8274..9137
/gene="yojK"
/function="unknown"
/note="similar to Bacillus subtilis Ydhe protein and to
Streptomyces lividans macrolide glycosyl transferase
encoded by the sequences presented in GenBank Accession
Numbers D88802 and M74717 respectively"
/codon_start=1
/transl_table=11
/product="YoJK"
/protein_id="AAC17859.1"
/db_xref="GI:3169327"
/translacion="MLKLPRFSLCTTFAMNEEFKEMMGAYMKGSLEDSPHYESYQQL"
Query Match 1.18; Score 21; DB 1; Length 15389;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 749 CAACATACACAGTTGCATATG 769
|||||
Db 9781 CAACATACACAGTTGCATATG 9801
|||||
RESULT 13
CBRG44E08/c 41830 bp DNA linear INV 04-NOV-2000
LOCUS Caenorhabditis briggsae cosmid G44E08, complete sequence.
DEFINITION AC084613
AC084613
AC084613.1 GI:11095062
VERSION HTG.
KEYWORDS
SOURCE Caenorhabditis briggsae.
ORGANISM Caenorhabditis briggsae.
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Rhabditoides; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE
1 (bases 1 to 41830)
The sequence of C. briggsae cosmid G44E08
AUTHORS Washington University Genome Sequencing Center.
TITLE The C. briggsae Genome Sequencing Project
JOURNAL Unpublished
REFERENCE
2 (bases 1 to 41830)
Wilson, R.
AUTHORS The sequence of C. briggsae cosmid G44E08
TITLE Unpublished
JOURNAL Unpublished
REFERENCE
3 (bases 1 to 41830)
Waterston, R.
AUTHORS Direct Submission
TITLE Submitted (04-NOV-2000) Department of Genetics, Washington
JOURNAL University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

```

COMMENT

Submitted by:

Genome Sequencing Center
Department of Genetics, Washington University,
St. Louis, MO 63110, USA
e-mail: jsplieth@wustl.edu

FEATURES

source

NOTICE: This sequence may not be the entire insert of this clone.
It may be shorter because we only sequence overlapping sections
once, or longer because we provide a small overlap between
neighboring submissions.
Location/Qualifiers
1..41830
/organism="Caenorhabditis briggsae"
/strain="GujArat G16"
/db_xref="taxon:6238"
/clone="G4E08"

BASE COUNT 12651 a 8319 c 7746 g 13114 t
ORIGIN

Query Match 1.1%; Score 21; DB 3; Length 41830;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1651 TTTTGCAATTATATGACCAT 1671
Db 6982 TTTTGCAATTATATGACCAT 6962

RESULT 14
AL355552/c 90157 bp DNA linear PRI 16-FEB-2001
LOCUS Human DNA sequence from clone RP11-217F17 on chromosome 13 contains
DEFINITION a ribosomal protein L7 pseudogene, STS and GSS, complete
sequence.

ACCESSION AL355552
VERSION AL355552.8 GI:11121041
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE
AUTHORS Direct Submission
TITLE Submitted (09-FEB-2001) Sanger Centre, Hinxton, Cambridgeshire,
JOURNAL CB10 ISA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Nov 8, 2000 this sequence version replaced gi:11062390.

During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.

The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 13, constructed by the Sanger Centre Chromosome 13
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr13

IMPORTANT: This sequence is not the entire insert of clone
RP11-217F17. It may be shorter because we sequence overlapping
sections only once, except for a 100 base overlap.
The true right end of clone RP11-217F17 is at 90157 in this
sequence. The true left end of clone RP11-562B17 is at 56085 in
this sequence. The true right end of clone RP11-155N20 is at 100 in
this sequence. This sequence was finished as follows unless
otherwise noted: all regions were either double-stranded or
sequenced with an alternate chemistry or covered by high quality

FEATURES

source

data (i.e., phred quality >= 30); an attempt was made to resolve
all sequencing problems, such as compressions and repeats; all
regions were covered by at least one plasmid subclone or more than
one M13 subclone; and the assembly was confirmed by restriction
digest. RP11-217F17 is from the library RP01-11.1 constructed by
the group of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBACe3.6.

Location/Qualifiers
1..90157
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="13"
/clone="RP11-217F17"
/clone_lib="RPCI-11.1"
complement(1..86)
/note="match: GSS: Em:AQ386890"
complement(1147..1651)
/note="match: GSS: Em:AQ522917"
2147..2490
/note="THE1B repeat: matches 1..364 of consensus"
2571..2754
/note="MER5A repeat: matches 2..186 of consensus"
2827..2892
/note="MER5A repeat: matches 118..186 of consensus"
4341..4631
/note="L2 repeat: matches 690..988 of consensus"
4630..5169
/note="L2 repeat: matches 2165..2750 of consensus"
5195..5510
/note="ALUSx repeat: matches 1..312 of consensus"
6835..7324
/note="match: GSS: Em:AQ433386"
7814..8054
/note="LIME3 repeat: matches 5792..6037 of consensus"
8057..8256
/note="LTRA1 repeat: matches 5..194 of consensus"
8257..8284
/note="L4 copies 2 mer aa 89% conserved"
8511..8644
/note="LTRA1 repeat: matches 617..741 of consensus"
8645..9755
/note="LIMEC repeat: matches 273..1383 of consensus"
9756..10048
/note="ALUSx repeat: matches 1..293 of consensus"
10049..10590
/note="LIMEC repeat: matches 183..1927 of consensus"
10591..10903
/note="ALUB repeat: matches 1..311 of consensus"
10904..11055
/note="LIMEC repeat: matches 1927..2078 of consensus"
12018..12331
/note="ALUSx repeat: matches 1..312 of consensus"
15295..15893
/note="LIME2 repeat: matches 19..572 of consensus"
16122..16719
/note="MER41B repeat: matches 1..606 of consensus"
16720..16958
/note="LIPR16 repeat: matches 5944..6165 of consensus"
17093..17373
/note="L2 repeat: matches 1539..1818 of consensus"
17376..17653
/note="MULTB repeat: matches 117..389 of consensus"
17872..18195
/note="ALUB repeat: matches 1..288 of consensus"
18202..18501
/note="ALUSx repeat: matches 1..303 of consensus"
19639..19678
/note="20 copies 2 mer tg 95% conserved"
19645..19676
/note="8 copies 4 mer tctg 100% conserved"
19680..19731
/note="13 copies 4 mer atat 73% conserved"

repeat_region 19758..19976
/note="AluDb repeat: matches 102. .287 of consensus"
repeat_region 20823..20866
/note="11 copies 4 mer tcct 79% conserved"
repeat_region 20983..21110
/note="64 copies 2 mer ta 90% conserved"
repeat_region 21326..21616
/note="MER46B repeat: matches 1. .236 of consensus"
repeat_region 21753..21930
/note="MIR repeat: matches 31. .207 of consensus"
repeat_region 22354..22551
/note="MER2 repeat: matches 3. .210 of consensus"
repeat_region 23293..23566
/note="AluY repeat: matches 39. .311 of consensus"
repeat_region 25962..26019
/note="2 copies 29 mer 96% conserved"
repeat_region 27578..27617
/note="10 copies 4 mer agag 90% conserved"
repeat_region 29695..29930
/note="MIR repeat: matches 20. .245 of consensus"
repeat_region 31437..31583
/note="MLT1-INTERNAL repeat: matches 6. .150 of consensus"
repeat_region 31584..32041
/note="L1P29 repeat: matches 5703. .6159 of consensus"
repeat_region 32042..32416
/note="MLT1-INTERNAL repeat: matches 150. .639 of consensus"
repeat_region 33079..33102
/note="12 copies 2 mer tt 100% conserved"
repeat_region 33488..33899
/note="MLT1B repeat: matches 1. .384 of consensus"
repeat_region 34181..34229
/note="MIR repeat: matches 21. .71 of consensus"
repeat_region 35227..35406
/note="2 copies 90 mer 85% conserved"
repeat_region 36236..36387
/note="MIR repeat: matches 20. .174 of consensus"
misc_feature Complement(37604..38241)
/note="match: GSS: Em:AQ357261"
misc_feature 37654..38484
/note="match: GSS: Em:AQ750273"
repeat_region 39395..39472
/note="39 copies 2 mer tt 64% conserved"
repeat_region 40268..40583
/note="AluX repeat: matches 1. .312 of consensus"
repeat_region 40687..40850
/note="MERS8D repeat: matches 1. .212 of consensus"
repeat_region 40881..40914
/note="MERS8D repeat: matches 164. .197 of consensus"
repeat_region 41288..41553
/note="MLT1E repeat: matches 307. .568 of consensus"
repeat_region 41582..41751
/note="MERS8D repeat: matches 269. .431 of consensus"
misc_feature Complement(42592..43084)
/note="match: STS: Em:G56777
/note="match: GSS: Em:AQ347427"
repeat_region 44442..44948
/note="L2 repeat: matches 2151. .2690 of consensus"
repeat_region 45948..46081
/note="MIR repeat: matches 38. .172 of consensus"
repeat_region 46286..46468
/note="MIR repeat: matches 48. .262 of consensus"
misc_feature Complement(47295..47695)
/note="match: GSS: Em:AQ347277
/note="match: STS: Em:G57525"
misc_feature 47698..48271
/note="match: GSS: Em:AQ779313"
misc_feature 48610..48816
/note="match: GSS: Em:B54651"
misc_feature Complement(48612..48788)
/note="match: GSS: Em:B51274"
misc_feature 48647..49389
/note="BA217F17.1"

CDS 48647..49389
/gene="BA217F17.1"
/note="bA217F17.1 (ribosomal protein L7 pseudogene)
match: proteins: Sw:Q12213 Sw:P25457 Tr:Q92U42 Sw:P11874
Sw:P17937 Sw:Q42208 Sw:P05737 Tr:Q92NU7"
/pseudo
/codon_start=1
/evidence=not_experimental
misc_feature complement(48659..48765)
/note="match: GSS: Em:AQ057230"
repeat_region 49847..50186
/note="MER6 repeat: matches 54. .399 of consensus"
Query Match 1.1%; Score 21; DB 9; Length 90157;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 537 TTAAGAAATCTGTAGAAAAG 557
|||||
Db 43199 TTAAGAAATCTGTAGAAAAG 43179
|||||
RESULT 15
AC000066 AC000066 94960 bp DNA linear PRI 21-DEC-1999
LOCUS Homo sapiens BAC clone CTA-293F11 from 7q21-7q22, complete
DEFINITION sequence.
AC000066 AC000066
VERSION AC000066.1 GI:3645943
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 94960)
AUTHORS Sulston, J.E. and Waterston, R.
TITLE Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE 99063792
PUBMED 9847074
REFERENCE 2 (bases 1 to 94960)
AUTHORS Wu, X., Graves, T. and Bradshaw, H.
TITLE The sequence of Homo sapiens BAC clone CTA-293F11
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 94960)
AUTHORS Waterston, R.
TITLE Direct Submission
JOURNAL Submitted (13-NOV-1996) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE 4 (bases 1 to 94960)
AUTHORS Waterston, R.
TITLE Direct Submission
JOURNAL Submitted (24-SEP-1998) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE 5 (bases 1 to 94960)
AUTHORS Waterston, R.
TITLE Direct Submission
JOURNAL Submitted (17-FEB-1999) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE 6 (bases 1 to 94960)
AUTHORS Waterston, R.
TITLE Direct Submission
JOURNAL Submitted (21-DEC-1999) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT On Sep 24, 1998 this sequence version replaced gi:1669365.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc>
Contact: sapiens@watson.wustl.edu
----- Summary Statistics
----- Center project name: H_RG293F11

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/GRB/CHR7> or <mailto:egreen@nhgri.nih.gov>

SOURCE INFORMATION:
Clone CTA-293F11 is from a release of the human BAC library CITR-HS-A. The library contains cloned DNA from human sperm. See: Shizuya et al., Proc. Natl. Acad. Sci. USA 89:8794-7 (1992); U-J. Kim et al., Genomics 34:213-8 (1996). The clone is available from Research Genetics, Inc. (<http://www.resgen.com>).
VECTOR: pBelobAC11
Selection: chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is CTB-104F4, 200 bp overlap; the clone sequenced to the right is GS1-541B18, 200 bp overlap. Actual start of this clone is unknown; actual end is at 94960 of RG293F11.

Location/Qualifiers

1. 94960

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="7"

/map="7q21-7q22"

/clone="CTA-293F11"

/clone_lib="CITR-HS-A"

2. 96

/rpt_family="MER1_type"

883. 990

/rpt_family="L1"

991. 1234

/rpt_family="Alu"

1421. 1722

/rpt_family="Alu"

2303. 2340

/note="match to EST AA329389 (NID:g1981781)"

2343. 2630

/note="match to EST AA329389 (NID:g1981781)"

4161. 4176

/rpt_family="Alu"

4177. 4203

/rpt_family="(TAA)n"

4204. 4480

/rpt_family="Alu"

4605. 4871

/rpt_family="Alu"

4882. 4935

/rpt_family="L2"

5906. 6100

/rpt_family="L2"

6316. 6367

/rpt_family="MIR"

6368. 6391

/rpt_family="L2"

6469. 6779

/rpt_family="Alu"

6865. 7757

/rpt_family="Alu"
complement(6037..8323)
/db_xref="GI:940344"
8193..8215
/rpt_family="AT_rich"
8612..8641
/rpt_family="AT_rich"
8785..8830
/rpt_family="AT_rich"
8833..9119
/rpt_family="(CA)n"
9792..9907
/rpt_family="Alu"
10207..11339
/rpt_family="L1"
11362..11489
/rpt_family="Alu"
11494..11762
/rpt_family="L1"
11764..11907
/rpt_family="L1"
11908..13266
/rpt_family="L1"
13267..13545
/rpt_family="Alu"
13546..14169
/rpt_family="Alu"
14170..14490
/rpt_family="L1"
14491..16242
/rpt_family="Alu"
16395..16858
/rpt_family="L1"
16859..17170
/rpt_family="L1"
17171..17233
/rpt_family="Alu"
17691..17989
/rpt_family="L1"
17996..18072
/rpt_family="L1"
18181..18249
/rpt_family="Mariner"
18401..18692
/rpt_family="L1"
18401..18692
/rpt_family="Alu"
19022..26200
/gene="WUSC:H_RG293F11.1"
join(<19022..19270,25591..25635,37460..37513,38187..38357,39923..40078,40905..41102,46150..48537,57731..57944,59551..59630,61431..61569,62319..62404,67540..67654,68116..68311,75197..75293,76814..76906,83721..84074,85976..>86200)
/gene="WUSC:H_RG293F11.1"
/note="match to AF026245 (PID:g2623068); but note differences at carboxy terminus; H_RG293F11.1"
/codon_start=1
/product="yotiao"
/protein_id="AAC60380.1"
/db_xref="GI:3645944"
/translation="PROKKAQSDGSPKQKKRKTSSKHDVSAHHDINDOSCN EMYNLSGVSTVPESTIMRTLSGHTSBGFSVLESEISTTDDCSSENGC SFVMTGKPTNLIREBERGVDDSYSEQAQDSPTLHMESEFLACKQHEIEINRELE EMRVYGTGEGUQLOEFPAAIKQDGIITQULANUQAARREDETFREFFLETSQK LQIQOQASSETLANSSTADLLQAKQIILHQOQLEQDHLDEYQKKEDFT MQISFLQEKIYVEMEDQKVENSKIEIOKEKTIIEBLNTRKIIIEEKTELKXKLT TADKLGLQLOIYQKNOEIKMKLEITNSKQSESEIKQMGVGEVLEKQKADT SQEETDIYVMEQETORLRLBRLDELMDGYQOIVYQKQELIRHMAEMMKTRHKG EMENLRYSNITNEDQIKANVAINEINIKLQDTQCKEKUKXELIIEBKALQ RQLEDVLELSFSREIQAROTTAQESKINAEKSLSTVEDLVAETVASESKEL EKHAEAVNTYKIKLEMLEKKNVLDMASSQAELETRTQULFSHEBELSKLED LEIHRINIEKIKNDIGIHYQOIQDGLQNESSQKIEITQFEDNLIITRONDLIEISK LKDLQSLVNSKSEMTLOINLEQKEIIRIOBEKEKGTLEBOVELOKTELEKQV KEXENDQEKPAQLEAENSILKDEKTELEDLKHTHPVSEERLIFLDSISKSDSV

Query Match 1.1%; Score 21; DB 9; Length 94960;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1395 ATTCTTTTGAAGAAATTC 1415
|||||
Db 81286 ATTCTTTTGAAGAAATTC 81306

RESULT 16
AL161743/c

LOCUS
DEFINITION Human DNA sequence from clone RP4-730B13 on chromosome 1q42.11-42.3
STSS and GSSs, complete sequence.
ACCESSION AL161743
VERSION AL161743.20 GI:11121016
KEYWORDS HTG; DISC1; schizophrenia.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (base 1 to 100170)
Direct Submission
Submitted (16-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
Requests: clonesrequest@sanger.ac.uk
On Nov 8, 2000 this sequence version replaced gi:10715861.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WormPEP; Information
on the WormPEP database can be found at
http://www.sanger.ac.uk/projects/C_elegans/wormep/
This sequence was generated from part of bacterial clone contigs of human
chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
Group. Further information can be found at
<http://www.sanger.ac.uk/HGP/Chri1>
This sequence is the entire insert of clone RP4-730B13 This
region was finished as follows unless otherwise noted: all
chemistry or covered by high quality data (i.e., phred quality >= 30);
an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. RP4-730B13 is from
the library RPLI-4 constructed by the group of Pieter de Jong. For
further details see
<http://www.chori.org/bacpac/home.htm>
VECTOR: pCYPAC2.

FEATURES
source
1..100170
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/map="g42.11-42.3"
/clone="RP4-730B13"
/clone_lib="RPLI-4"
complement(join(137..255,17641..17905,67802..67862))
/genes="DISC1"

gene
complement(join(<137..255,17641..17905,67802..>67862))

CDS
complement(join(<137..255,17641..17905,67802..>67862))

```

repeat_region 21549..21786
/note="119 copies 2 mer ag 56% conserved"
repeat_region 21828..21951
/note="L2 repeat: matches 966..1088 of consensus"
repeat_region 21952..22071
/note="L1MC/D repeat: matches 5371..5496 of consensus"
repeat_region 22107..22677
/note="L1M4 repeat: matches 2922..3512 of consensus"
repeat_region 22825..23558
/note="L1M4 repeat: matches 2131..2925 of consensus"
repeat_region 23635..25145
/note="L1MD2 repeat: matches 4794..6307 of consensus"
repeat_region 25313..25611
/note="ALUSX repeat: matches 2..302 of consensus"
repeat_region 25800..26697
/note="L1MD repeat: matches 979..1932 of consensus"
repeat_region 26792..26881
/note="16 copies 9 mer tggtttgc 71% conserved"
repeat_region 26799..26878
/note="40 copies 2 mer tc 63% conserved"
repeat_region 27364..27547
/note="L1MD repeat: matches -17..167 of consensus"
/note="complement(27754..28002)"
misc_feature /gene="DISC1"
/note="match: GSS: Em:AQ019129"
repeat_region 28092..28203
/note="HML1 repeat: matches 626..730 of consensus"
repeat_region 30215..31339
/note="L1PA2 repeat: matches 5020..6144 of consensus"
repeat_region 31342..31377
/note="18 copies 2 mer aa 83% conserved"
repeat_region 31979..32041
/note="7 copies 9 mer tcatcatca 73% conserved"
repeat_region 32052..32121
/note="MTR repeat: matches 158..227 of consensus"
repeat_region 32137..32407
/note="MTR repeat: matches 1..261 of consensus"
repeat_region 32979..33095
/note="L2 repeat: matches 2593..2705 of consensus"
repeat_region 34127..34397
/note="L1ME repeat: matches 5108..5386 of consensus"
repeat_region 34495..34664
/note="L1ME repeat: matches 5567..5747 of consensus"
repeat_region 36753..36960
/note="MIR repeat: matches 35..262 of consensus"
repeat_region 37937..38298
/note="MTR1A1 repeat: matches 1..365 of consensus"
repeat_region 38332..38471
/note="MER5A repeat: matches 9..173 of consensus"
repeat_region 38492..38783
/note="ALUSX repeat: matches 1..295 of consensus"
repeat_region 39472..39939
/note="match: GSS: Em:AQ068363"
misc_feature 39501..39917
/note="match: GSS: Em:AQ088543"
/note="complement(39885..40383)"
misc_feature /gene="DISC1"
/note="match: GSS: Em:AQ121406"
repeat_region 40529..40699
/note="MIR repeat: matches 59..233 of consensus"
repeat_region 41234..41325
/note="MIR repeat: matches 166..252 of consensus"
repeat_region 41326..41651
/note="ALUSG repeat: matches 1..311 of consensus"
repeat_region 41652..41792
/note="MIR repeat: matches 7..166 of consensus"
repeat_region 41864..42129
/note="L2 repeat: matches 2408..2688 of consensus"
repeat_region 42614..42791
/note="MIR repeat: matches 32..211 of consensus"
repeat_region 42877..43020

```

1.1%; Score 21; DB 9; Length 100170;

Best Local Similarity 100.0%; Pred. No. 10;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1400 TTTTGGAGAAATTCATA 1420
DB 21386 TTTTGGAGAAATTCATA 21366

RESULT 17
AC122235_2/c
WPCOMMENT
Sequence split into 5 fragments LOCUS AC122235 Accession AC122235

Fragment Name	Begin	End
AC122235_0	1	110000
AC122235_1	100001	210000
AC122235_2	200001	310000
AC122235_3	300001	410000
AC122235_4	400001	434777

Continuation (3 of 5) of AC122235 from base 200001 (AC122235 Mus musculus chromosome UR

Query Match 1.1%; Score 21; DB 2; Length 110000;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 992 CTACAGAACTGCTTCT 1012
DB 83906 CTACAGAACTGCTTCT 83886

RESULT 18
AC124954/c
LOCUS 114741 bp DNA linear HTG 15-AUG-2002
DEFINITION Medicago truncatula clone mth1-10g9, WORKING DRAFT SEQUENCE, 5
ORDERED pieces.
AC124954
AC124954.5 GI:21717195
VERSION HTG: HTGS PHASE2; HTGS_DRAFT.
KEYWORDS barrel medic.
SOURCE Medicago truncatula
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifoliaceae;
Medicago.

REFERENCE 1 (bases 1 to 114741)
AUTHORS Shauli,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
Cook,D., Kim,D., and Roe,B.A.

TITLE Medicago truncatula BAC Clone mth1-10g9
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 114741)
Shauli,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
Cook,D., Kim,D., and Roe,B.A.

TITLE Direct Submission
JOURNAL Submitted (20-JUN-2002) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA

REFERENCE 3 (bases 1 to 114741)
AUTHORS Shauli,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
Cook,D., Kim,D., and Roe,B.A.

TITLE Direct Submission
JOURNAL Submitted (15-AUG-2002) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA

COMMENT ON Jul 9, 2002 this sequence version replaced gi:21672150.
----- Genome Center
Center: Department Of Chemistry And Biochemistry
The University Of Oklahoma
Center code:UOKNOR

* NOTE: This is a 'working draft' sequence. It currently
* consists of 5 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have

* provided by the submitter.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.

* 1 3389: contig of 3389 bp in length
 * 3390 3489: gap of unknown length
 * 3490 19779: contig of 16290 bp in length
 * 19780 19879: gap of unknown length
 * 19880 39080: contig of 19201 bp in length
 * 39081 39180: gap of unknown length
 * 39181 67046: contig of 27866 bp in length
 * 67047 67146: gap of unknown length
 * 67147 114741: contig of 47595 bp in length.

FEATURES

source

1. 114741
 /organism="Medicago truncatula"
 /db xref="taxon:3880"
 /clone="mchl-1093"
 /clone.lib="Medicago truncatula BAC library H1"
 /clone.lib="Medicago truncatula BAC library H1"
 /clone.lib="Medicago truncatula BAC library H1"

BASE COUNT 39037 a 19145 c 18697 g 37461 t 401 others

ORIGIN

Query Match 1.1%; Score 21; DB 2; Length 114741;
 Best Local Similarity 100.0%; Pred. No. 10;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1408 GAAATTCAAATAGATTGCA 1428

|||||
 Db 62328 GAAATTCAAATAGATTGCA 62308

RESULT 19

AC129991

LOCUS

DEFINITION Rattus norvegicus clone CH230-101023, *** SEQUENCING IN PROGRESS

*** 50 unordered pieces.

AC129991

VERSION

AC129991.1

GI:22138135

HTG; HTGS PHASE1.

KEYWORDS

SOURCE

Norway rat.

ORGANISM

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

1 (bases 1 to 123091)

Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,

Alsbrooks,S.L., Amarantunge,H.C., Are,J.R., Ayele,M., Banks,T.,

Barbaria,J., Benton,J., Binage,K., Blankenburg,K., Bonnin,D.,

Boucha,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,

Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,

Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,

Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,

Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,

Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,

Delaney,K.R., Deigado,O., Denu,A.L., Ding,Y., Dinh,H.H.,

Dellaway,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,

Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,

Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,

Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,

Garrell,J.H., Guevara,W., Guarnatone,P., Hale,S., Hamilton,K.,

Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,

Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,

Honsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,

Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,

Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,

Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,

Li,J., Li,Z., Lichkarge,O., Lieu,C., Liu,J., Liu,W., Loulised,H.,

Lozad,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,

Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,

Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,

Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,

Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,

Nguyen,N., Nickerson,E., Nwokenkwo,S., Ogun,M., Okwuonu,G.,

Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
 Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
 Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
 Scherer,S., Scott,G., Shen,H., Shoohtari,N., Sisson,I.,
 Sodergren,E., Sonaik,T., Sparks,A., Stanley,H., Stone,H.,
 Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
 Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
 Usmani,K., Vasquez,L., Vera,V., Villaion,D., Vinson,R., Wang,Q.,
 Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
 Williams,G., Williamson,A., Wlezyk,R., Wooden,S., Worley,K.,
 Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
 Weinstock,G., and Gibbs,R.

COMMENT

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: KAOQ

Center clone name: CH230-101023

----- Summary Statistics

Sequencing vector: Plasmid;

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 67169 bases at least Q40

Consensus quality: 71508 bases at least Q30

Consensus quality: 75279 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).

* NOTE: This is a 'working draft' sequence. It currently

* consists of 60 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

* 1 1092: contig of 1092 bp in length

* 1093 1152: gap of unknown length

* 1193 2232: contig of 1040 bp in length

* 2233 2332: gap of unknown length

* 2333 3881: contig of 1549 bp in length

* 3882 3981: gap of unknown length

* 3982 5083: contig of 1102 bp in length

* 5084 5183: gap of unknown length

* 5184 6592: contig of 1309 bp in length

* 6593 7767: contig of 1175 bp in length

* 7768 8991: contig of 1124 bp in length

* 8992 9091: gap of unknown length

* 9092 10558: contig of 1467 bp in length

* 10559 10658: gap of unknown length

* 10659 12099: contig of 1441 bp in length

* 12100 12199: gap of unknown length

* 12200 13294: contig of 1095 bp in length

* 13295 13394: gap of unknown length

* 13395 14659: contig of 1265 bp in length

* 14660 14759: gap of unknown length

* 14760 15994: contig of 1235 bp in length

* 15995 16094: gap of unknown length

* 16095 17255: contig of 1161 bp in length

* 17256 17355: gap of unknown length

* 17356 18591: contig of 1236 bp in length

18592 18691: gap of unknown length
18692 20268: contig of 1577 bp in length
20269 20368: gap of unknown length
20369 21874: contig of 1506 bp in length
21875 21974: gap of unknown length
21975 23378: contig of 1404 bp in length
23379 23478: gap of unknown length
23479 24606: contig of 1128 bp in length
24607 24706: gap of unknown length
24707 26033: contig of 1327 bp in length
26034 27532: gap of unknown length
27533 27632: contig of 1399 bp in length
27633 29257: gap of unknown length
29258 29357: gap of unknown length
29358 30896: contig of 1539 bp in length
30897 32545: gap of unknown length
32546 32645: contig of 1549 bp in length
32646 33774: gap of unknown length
33775 33874: contig of 1129 bp in length
33875 35039: gap of unknown length
35039 35139: contig of 1215 bp in length
35139 37444: gap of unknown length
37445 37544: contig of 2255 bp in length
37545 39032: gap of unknown length
39032 39132: contig of 1538 bp in length
39132 40532: gap of unknown length
40533 40632: contig of 1350 bp in length
40633 41737: gap of unknown length
41737 41837: contig of 1075 bp in length
41837 43935: gap of unknown length
43935 44035: contig of 2188 bp in length
44035 44396: gap of unknown length
44396 45448: contig of 1253 bp in length
45448 47732: gap of unknown length
47732 47832: contig of 2344 bp in length
47832 49418: gap of unknown length
49418 49518: contig of 1526 bp in length
49518 51246: gap of unknown length
51246 51346: contig of 1728 bp in length
51346 52750: gap of unknown length
52750 52850: contig of 1404 bp in length
52850 54709: gap of unknown length
54709 54809: contig of 1859 bp in length
54809 56706: gap of unknown length
56706 56806: contig of 1897 bp in length
56806 59494: gap of unknown length
59494 59594: contig of 2688 bp in length
59594 60822: gap of unknown length
60822 60922: contig of 1228 bp in length
60922 62056: gap of unknown length
62056 62156: contig of 1134 bp in length
62156 64918: gap of unknown length
64918 65038: contig of 2782 bp in length
65038 66239: gap of unknown length
66239 66339: contig of 1261 bp in length
66339 68146: gap of unknown length
68146 68246: contig of 1747 bp in length
68246 69841: gap of unknown length
69841 69941: contig of 1595 bp in length
69941 71777: gap of unknown length
71777 71877: contig of 1836 bp in length
71877 73269: gap of unknown length
73269 73369: contig of 1392 bp in length
73369 75564: gap of unknown length
75564 75664: contig of 2195 bp in length
75664 77306: gap of unknown length
77306 77406: contig of 1642 bp in length
77406 79833: gap of unknown length
79833 79934: contig of 2427 bp in length
79934 82424: gap of unknown length
82424 82524: contig of 2491 bp in length
82524 gap of unknown length

Query Match 1.1%; Score 21; DB 2; Length 123091;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1403 TTGAAGAAATTCATATAGAT 1423
Db 47571 TTGAAGAAATTCATATAGAT 47591
RESULT 20
AC127017/c
LOCUS
AC127017 Rattus norvegicus clone CH230-66a17, *** SEQUENCING IN PROGRESS
DEFINITION
AC127017
ACCESSION
AC127017.1 GI:21734917
VERSION
AC127017.1
KEYWORDS
HTG; HTGS-PHASE1.
SOURCE
Norway rat.
ORGANISM
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
1 (bases 1 to 124028)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaraltinge,H.C., Are,J.R., Ayale,M., Banks,T.,
Barbata,J., Benson,J., Bimaga,K., Blankenburg,K., Bonnin,D.,
Bouk,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,T., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Dublin,K.J.,
Einhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escoto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,U., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Haylak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hogue,M., Holloway,C., Hollins,B.,
Homai,F., Howard,S., Huber,J., Hulix,S., Hume,D., Jackson,L.B.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudan,S.,
Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Kratovic,J., Kureishi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louisedge,H.,
Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Mawhinney,B., McLeod,M.P., Meador,M., Mel,G., Metzger,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Moser,M., Neal,D., Newton,J., Newton,S., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okunolu,G.,
Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Pickens,R., Primus,E., Pu,L.L., Ruiz,M., Ren,Y.,
Rives,M., Rojas,A., RojudoKan,I., Rolfe,M., Ruiz,S., Savery,G.,
Scherer,S., Scott,G., Shen,H., Shoostari,N., Sisson,T.,
Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,

Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
Williams,C., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.
Direct Submision
Unpublished
2 (bases 1 to 124028)
Worley,K.C.
Direct Submision
Submitted (12-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 124028)
Worley,K.C.
Direct Submision
Submitted (31-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
----- Genome Center of Medicine
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: RAWX
Center clone name: CH230-66A17
----- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100k of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 66056 bases at least Q40
Consensus quality: 71270 bases at least Q30
Consensus quality: 74172 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 69 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 1003: contig of 1003 bp in length
* 1004 1103: gap of unknown length
* 1104 2256: contig of 1153 bp in length
* 2257 2356: gap of unknown length
* 2357 3434: contig of 1078 bp in length
* 3435 3534: gap of unknown length
* 3535 4606: contig of 1072 bp in length
* 4607 4706: gap of unknown length
* 4707 5786: contig of 1080 bp in length
* 5787 5886: gap of unknown length
* 5887 7370: contig of 1484 bp in length
* 7371 7470: gap of unknown length
* 7471 8841: contig of 1371 bp in length
* 8842 8941: gap of unknown length
* 8942 10117: contig of 1176 bp in length
* 10118 10217: gap of unknown length
* 10218 11272: contig of 1055 bp in length
* 11273 11372: gap of unknown length
* 11373 12418: contig of 1046 bp in length
* 12419 12518: gap of unknown length
* 12519 14138: contig of 1620 bp in length
* 14139 14238: gap of unknown length
* 14239 15243: contig of 1005 bp in length
* 15244 15343: gap of unknown length
* 15344 16432: contig of 1089 bp in length
* 16433 16532: gap of unknown length
* 16533 17608: contig of 1076 bp in length
* 17609 17708: gap of unknown length
*
* 17709 18850: contig of 1141 bp in length
* 18850 18949: gap of unknown length
* 18949 20015: contig of 1066 bp in length
* 20016 20116: gap of unknown length
* 20117 21605: contig of 1490 bp in length
* 21606 21705: gap of unknown length
* 21706 23231: contig of 1526 bp in length
* 23232 23331: gap of unknown length
* 23332 24853: contig of 1522 bp in length
* 24854 24953: gap of unknown length
* 24954 26081: contig of 1128 bp in length
* 26082 26181: gap of unknown length
* 26182 27398: contig of 1217 bp in length
* 27399 27498: gap of unknown length
* 27499 28574: contig of 1076 bp in length
* 28575 28674: gap of unknown length
* 28675 29750: contig of 1076 bp in length
* 29751 29850: gap of unknown length
* 29851 31004: contig of 1154 bp in length
* 31005 31104: gap of unknown length
* 31105 32493: contig of 1389 bp in length
* 32494 32593: gap of unknown length
* 32594 33769: contig of 1176 bp in length
* 33770 33869: gap of unknown length
* 33870 35109: contig of 1240 bp in length
* 35110 35209: gap of unknown length
* 35210 36997: contig of 1788 bp in length
* 36998 37097: gap of unknown length
* 37098 38206: contig of 1109 bp in length
* 38207 38306: gap of unknown length
* 38307 39463: contig of 1157 bp in length
* 39464 39563: gap of unknown length
* 39564 41966: contig of 2403 bp in length
* 41967 42066: gap of unknown length
* 42067 43357: contig of 1291 bp in length
* 43358 43457: gap of unknown length
* 43458 45525: contig of 2068 bp in length
* 45526 45626: gap of unknown length
* 45627 47806: contig of 2181 bp in length
* 47807 47906: gap of unknown length
* 47907 49278: contig of 1373 bp in length
* 49279 49379: gap of unknown length
* 49380 51108: contig of 1729 bp in length
* 51109 51209: gap of unknown length
* 51210 52355: contig of 1147 bp in length
* 52356 52455: gap of unknown length
* 52456 54114: contig of 1659 bp in length
* 54115 54214: gap of unknown length
* 54215 55878: contig of 1664 bp in length
* 55879 55978: gap of unknown length
* 55979 57262: contig of 1284 bp in length
* 57263 57363: gap of unknown length
* 57364 59333: contig of 2371 bp in length
* 59334 59833: gap of unknown length
* 59834 62133: contig of 2300 bp in length
* 62134 62233: gap of unknown length
* 62234 63745: contig of 1512 bp in length
* 63746 63845: gap of unknown length
* 63846 65057: contig of 1212 bp in length
* 65058 65157: gap of unknown length
* 65158 66293: contig of 1136 bp in length
* 66294 66393: gap of unknown length
* 66394 68088: contig of 1695 bp in length
* 68089 68188: gap of unknown length
* 68189 69189: contig of 1001 bp in length
* 69190 69289: gap of unknown length
* 69290 70839: contig of 1550 bp in length
* 70840 72523: gap of unknown length
* 72524 72623: contig of 1584 bp in length
* 72624 74389: gap of unknown length
* 74390 74489: gap of unknown length
* 74490 75927: contig of 1438 bp in length

TITLE
JOURNAL
REFERENCE

AUTHORS
TITLE
JOURNAL

COMMENT

* 75928 76027: gap of unknown length
 * 76028 77665: contig of 1538 bp in length
 * 77666 77665: gap of unknown length
 * 77666 79731: contig of 2066 bp in length
 * 79732 79831: gap of unknown length
 * 79832 83121: contig of 3290 bp in length
 * 83122 83221: gap of unknown length
 Query Match 1.1%; Score 21; DB 2; Length 124028;
 Best Local Similarity 100.0%; Pred. No. 10;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1221 TAAGCAGCAACTATCGAAGA 1241
 |||||
 DB 45185 TAAGCAGCAACTATCGAAGA 45165

RESULT 21
 AL445200 133054 bp DNA linear HTG 10-JUL-2001
 LOCUS Homo sapiens chromosome 1 clone RP5-852F22, *** SEQUENCING IN
 DEFINITION PROGRESS *** 16 unordered pieces.
 ACCESSION AL445200
 VERSION AL445200.1 GI:10716452
 KEYWORDS HTG; HTGS_PHASE1; HTGS_CANCELLED.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE
 AUTHORS McIay, K.
 TITLE Direct Submission
 JOURNAL Submitted (09-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
 CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
 requests: clonerequest@sanger.ac.uk

COMMENT

----- Genome Center
 Center: Sanger Centre
 Center code: SC
 Web site: http://www.sanger.ac.uk
 Contact: humquery@sanger.ac.uk
 ----- Project Information
 Center project name: dj852F22
 ----- Summary Statistics
 Assembly program: XGAP4; version 4.5
 Sequencing vector: plasmid; L08752; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Consensus quality: 124206 bases at least Q40
 Consensus quality: 129947 bases at least Q30
 Consensus quality: 129947 bases at least Q20
 Insert size: 131534; sum-of-contigs
 Insert size: 149216; 5.4% error; agarose-fp
 Quality coverage: 3.44x in Q20 bases; sum-of-contigs Quality
 coverage: 3.12x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 16 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

* 1 6858: contig of 6858 bp in length
 * 6859 6958: gap of 100 bp
 * 6959 11344: contig of 4386 bp in length
 * 11345 11444: gap of 100 bp
 * 11445 15274: contig of 3830 bp in length
 * 15275 15374: gap of 100 bp
 * 15375 18330: contig of 2956 bp in length
 * 18331 18430: gap of 100 bp
 * 18431 38813: contig of 20383 bp in length
 * 38814 38913: gap of 100 bp
 * 38914 42991: contig of 4078 bp in length

FEATURES

source * 42992 43091: gap of 100 bp
 * 43092 46478: contig of 3387 bp in length
 * 46479 46578: gap of 100 bp
 * 46579 61624: contig of 15046 bp in length
 * 61625 61724: gap of 100 bp
 * 61725 70468: contig of 8744 bp in length
 * 70469 70568: gap of 100 bp
 * 70569 78278: contig of 7710 bp in length
 * 78279 78378: gap of 100 bp
 * 78379 80926: contig of 2648 bp in length
 * 80927 81026: gap of 100 bp
 * 81027 89376: contig of 8350 bp in length
 * 89377 89476: gap of 100 bp
 * 89477 92606: contig of 3130 bp in length
 * 92607 92706: gap of 100 bp
 * 92707 96333: contig of 3627 bp in length
 * 96334 96433: gap of 100 bp
 * 96434 114593: contig of 18160 bp in length
 * 114594 114694: gap of 100 bp
 * 114694 133054: contig of 18361 bp in length.
 Location/Qualifiers
 1..133054
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="1"
 /clone="RP5-852F22"
 /clone_id="RPC1-5"
 1..6858
 /note="assembly_fragment:01045
 fragment_chain:1
 clone_end:SP6
 vector_side:left"
 6959..11344
 /note="assembly_fragment:00181
 fragment_chain:1"
 11445..15274
 /note="assembly_fragment:01118
 fragment_chain:1"
 15375..18330
 /note="assembly_fragment:00832
 fragment_chain:1"
 18431..38813
 /note="assembly_fragment:00599
 fragment_chain:1"
 38914..42991
 /note="assembly_fragment:00760
 fragment_chain:1"
 43092..46478
 /note="assembly_fragment:00784
 fragment_chain:1"
 46579..61624
 /note="assembly_fragment:00618
 fragment_chain:2"
 61725..70468
 /note="assembly_fragment:01180
 fragment_chain:2"
 70569..78278
 /note="assembly_fragment:00949
 fragment_chain:2"
 78379..80926
 /note="assembly_fragment:00888
 fragment_chain:3"
 81027..89376
 /note="assembly_fragment:00772
 fragment_chain:3"
 89477..92606
 /note="assembly_fragment:00426
 fragment_chain:3"
 92707..96333
 /note="assembly_fragment:00451
 fragment_chain:4"
 96434..114593
 /note="assembly_fragment:00396


```

misc_feature      fragment_chain:4"
114694.._133054
/notes="assembly_fragment:00882
clone_end:T7
vector_side:right"
BASE COUNT      39306 a 27031 c 27151 g 38061 t 1502 others
ORIGIN
Query Match      1.1%; Score 21; DB 2; Length 133054;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1400 TTTTGTGAAGAAATTCAAATA 1420
|||||
Db 97814 TTTTGAAGAAATTCAAATA 97834

RESULT 22
AC114333/c
LOCUS            AC114333      145661 bp      DNA      linear      HTG 07-MAR-2002
DEFINITION      Canis familiaris clone RP81-588, WORKING DRAFT SEQUENCE, 4
unordered pieces.
ACCESSION       AC114333
VERSION         AC114333.1 GI:19224982
KEYWORDS        HTG; HTGS_P1; HTGS_DRAFT.
SOURCE          Canis familiaris
ORGANISM        Canis familiaris
REFERENCE       Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS        Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
1 (bases 1 to 145661)
Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S.,
Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P.,
Ho,S.-L., Idol,J.R., Karlins,E., Laric,P., Lee-Lin,S.-Q.,
Legaspi,R., Madiro,Q.L., Madiro,V.B., Masiello,C., Maskeri,B.,
Mastrian,S.D., McCloskey,J.C., McDowell,J., Pearson,R., Prasad,A.,
Stantrop,S., Thomas,J.W., Thomas,P.J., Touchman,J.W., Prasad,A.,
Tsurgeon,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L.,
Young,A., Zhang,L.-H. and Green,E.D.
NISC Comparative Sequencing Initiative
Unpublished
2 (bases 1 to 145661)
Green,E.D.
Direct Submission
Submitted (07-MAR-2002) NIH Intramural Sequencing Center, 8717
Government Circle, Gaithersburg, MD 20877, USA
----- Genome Center
Center: NIH Intramural Sequencing Center
Center code: NISC
Web site: http://www.nisc.nih.gov
Contact: nisc.mouse@nhgri.nih.gov
----- Project Information
Center project name: Civ
Center clone name: 005B08
----- Summary Statistics
Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 144074 bases at least Q40
Consensus quality: 144474 bases at least Q30
Consensus quality: 144791 bases at least Q20
Insert size: 120000; agarose-fp
Insert size: 145361; sum-of-ctnigs
Quality coverage: 9.82x in Q20 bases; agarose-fp
Quality coverage: 8.10x in Q20 bases; sum-of-ctnigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will

```

```

* be preserved.
* 1 6901: contig of 6901 bp in length
* 6902 7001: gap of unknown length
* 7002 45737: contig of 38736 bp in length
* 45738 45837: gap of unknown length
* 45838 90402: contig of 44565 bp in length
* 90403 90502: gap of unknown length
* 90503 145661: contig of 55159 bp in length.
FEATURES
source
1..145661
Location/Qualifiers
/organism="Canis familiaris"
/db_xref="taxon:9615"
/clone="RP81-588"
/clone_lib="RP81"
1..6901
note="assembly_fragment"
7002..45737
note="assembly_fragment"
45838..90402
note="assembly_fragment"
90503..145661
note="assembly_fragment"
BASE COUNT 43888 a 28373 c 28107 g 44993 t 300 others
ORIGIN
Query Match      1.1%; Score 21; DB 2; Length 145661;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1737 GTAAAGAAATCGCTCAAAAT 1757
|||||
Db 81634 GTAAAGAAATCGCTCAAAAT 81614

RESULT 23
AL356867
LOCUS            AL356867      147803 bp      DNA      linear      PRI 12-FEB-2002
DEFINITION      Human DNA sequence from clone RP1-20N18 on chromosome 1, complete
sequence.
ACCESSION       AL356867
VERSION         AL356867.23 GI:18655972
KEYWORDS        HTG.
SOURCE          human.
ORGANISM        Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 147803)
Direct Submission
Submitted (12-FEB-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquerry@sanger.ac.uk Clone requests: clonequest@sanger.ac.uk
On Feb 13, 2002 this sequence version replaced gi:16304699.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em:, EMBL; Sw:,
SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
database can be found at
http://www.sanger.ac.uk/projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping

```

Group. Further information can be found at
<http://www.sanger.ac.uk/HGP/Ch1>
 RPI-20N18 is from the library RPI-1 constructed by the group of
 Pieter de Jong. For further details see
<http://www.chori.crg/bacpac/home.htm>
 VECTOR: pCYPAC2

IMPORTANT: This sequence is not the entire insert of clone
 RPI-20N18 it may be shorter because we sequence overlapping
 sections only once, except for a short overlap.
 The true right end of clone RPI-20N18 is at 147803 in this
 sequence. The true left end of clone RPI-59H12 is at 107462 in this
 sequence. The true right end of clone RPI-13P20 is at 2000 in this
 sequence.

FEATURES
 source
 1. 147803
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="1"
 /clone="RPI-20N18"
 /clone_lib="RPI-1"

BASE COUNT 45512 a 30941 c 29246 g 42104 t
 ORIGIN

Query Match 1.1%; Score 21; DB 9; Length 147803;
 Best Local Similarity 100.0%; Pred. No. 10;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 39 CTGCTAATGCAATATTCG 59
 |||||
 Db 67488 CTGCTAATGCAATATTCG 67508

RESULT 24
 AC011024/c 157326 bp DNA linear HTG 26-MAY-2000
 LOCUS
 DEFINITION Homo sapiens clone RPI-115D12, WORKING DRAFT SEQUENCE, 6 unordered
 pieces.

AC011024
 AC011024.3 GI:7107790
 HTG; HTGS_PHASE1; HTGS_DRAFT.
 KEYWORDS
 SOURCE
 ORGANISM
 Homo sapiens

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 1 (bases 1 to 157326)
 Birren, B., Linton, L., Nussbaum, C. and Lander, E.
 Homo sapiens, clone RPI-115D12
 Unpublished
 2 (bases 1 to 157326)
 Birren, B., Linton, L., Nussbaum, C., Lander, E., Allen, N., Anderson, M.,
 Baldwin, J., Barina, N., Beckert, R., Boguslavsky, L., Boukhgalter, B.,
 Brown, A., Castle, A., Colangelo, M., Collins, S., Collymore, A.,
 Cooke, P., Dearrellano, K., Dewar, K., Domingo, M., Donelan, L., Doyle, M.,
 Ferreira, P., Fitzhugh, W., Forrest, C., Funke, R., Gage, D.,
 Galagan, J., Gardyna, S., Grant, G., Hagos, B., Heath, A., Horton, L.,
 Howland, J. C., Johnson, R., Jones, C., Kann, L., Karata, A., Klein, J.,
 Lehoczy, J., Lieu, C., Locke, K., MacDonald, P., Marquis, N.,
 McEwan, P., McGuirk, A., McKernan, K., McLaughlin, J., Meldrum, J.,
 Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,
 Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P.,
 Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
 Teste, S., Tifreil, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X.,
 Wyman, D., Ye, W. J., Zimmer, A. and Zody, M.

TITLE

JOURNAL

COMMENT

Submitted (30-SEP-1999) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Feb 28, 2000 this sequence version replaced gi:563609.
 All repeats were identified using RepeatMasker:
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>
 ----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: MIBR

Web site: <http://www.ssg.wi.mit.edu>
 Contact: sequence_submissions@genome.wi.mit.edu
 Project Information
 Center project name: L1988
 Center clone name: 115.D.12

----- Summary Statistics
 Sequencing vector: M13; M77815; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 142326 bases at least Q40
 Consensus quality: 151063 bases at least Q30
 Consensus quality: 154192 bases at least Q20
 Insert size: 150000; agarose-fp
 Insert size: 156826; sum-of-ctnigs
 Quality coverage: 5.4 in Q20 bases; agarose-fp
 Quality coverage: 5.2 in Q20 bases; sum-of-ctnigs

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 6 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 8117: contig of 8117 bp in length
 * 8118 8217: gap of 100 bp
 * 8218 24311: contig of 16094 bp in length
 * 24312 24411: gap of 100 bp
 * 24412 46273: contig of 21862 bp in length
 * 46274 46373: gap of 100 bp
 * 46374 77849: contig of 31476 bp in length
 * 77850 77949: gap of 100 bp
 * 77950 112313: contig of 34364 bp in length
 * 112314 112413: gap of 100 bp
 * 112414 157326: contig of 44913 bp in length.

FEATURES
 source
 1. 157326
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="RPI-115D12"
 /clone_lib="RPI-11 Human Male BAC"

misc_feature
 1. 8117
 /note="assembly_fragment"
 clone_end:17
 vector_side:right"
 misc_feature
 8218..24311
 /note="assembly_fragment"
 24412..46273
 /note="assembly_fragment"
 clone_end:SP6
 vector_side:left"
 misc_feature
 46374..77849
 /note="assembly_fragment"
 77950..112313
 /note="assembly_fragment"
 112414..157326
 /note="assembly_fragment"
 BASE COUNT 50401 a 28032 c 28903 g 49486 t 504 others
 ORIGIN

Query Match 1.1%; Score 21; DB 2; Length 157326;
 Best Local Similarity 100.0%; Pred. No. 10;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 537 TTAATAATCTGTAAGAAAG 557
 |||||
 Db 3618 TTAATAATCTGTAAGAAAG 3598

RESULT 25
 AC106409 168632 bp DNA linear HTG 17-JUL-2002
 LOCUS

DEFINITION

ACCESSION AC106409
VERSION AC106409.2 GI:21744154
KEYWORDS HTG; HTGS PHASE1.
SOURCE Norway rat.
ORGANISM Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE

AUTHORS

1 (bases 1 to 168632)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Albrooks,S.L., Amaraturunge,H.C., Ars,J.R., Ayele,M., Banks,T., Barbara,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.B., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,I., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Honsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,B., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., McLeod,M.P., Meador,M., Mei,G., Metzker,M., Massey,E., Mawhiney,E., Mitchell,I., Mohabbat,K., Morgan,M., Morris,S., Miner,G., Miner,Z., Mitchell,I., Newton,N., Nguyen,A., Nguyen,N., Moser,M., Neal,D., Newton,S., Nwokenwo,S., Oguh,M., Okwuonu,G., Nguyen,N., Nickerson,E., Nwokenwo,S., Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczek,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinstein,G. and Gibbs,R.

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

Submitted (12-JAN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE

AUTHORS

TITLE

JOURNAL

Submitted (17-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

COMMENT

On Jul 13, 2002 this sequence version replaced gi:18138930.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

----- Project name: GKVR

Center clone name: CH230-80K5

----- Summary Statistics

Sequencing vector: Plasmid;

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly: Phrap; version 0.990329

Consensus quality: 119664 bases at least Q40

Consensus quality: 125868 bases at least Q30

Consensus quality: 130623 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).

* NOTE: This is a 'working draft' sequence. It currently

* consists of 59 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

* -----

* 1 1501: contig of 1501 bp in length

* 1502: gap of unknown length

* 1602: contig of 1258 bp in length

* 2859: gap of unknown length

* 2959: gap of unknown length

* 2860: contig of 1193 bp in length

* 4132: gap of unknown length

* 4252: contig of 1229 bp in length

* 5481: gap of unknown length

* 5581: contig of 1554 bp in length

* 7135: gap of unknown length

* 7235: contig of 1210 bp in length

* 8445: gap of unknown length

* 8545: gap of unknown length

* 9852: contig of 1307 bp in length

* 9952: gap of unknown length

* 1061: contig of 1009 bp in length

* 11061: gap of unknown length

* 12260: contig of 1199 bp in length

* 12360: gap of unknown length

* 13736: contig of 1376 bp in length

* 13836: gap of unknown length

* 14939: contig of 1103 bp in length

* 15039: gap of unknown length

* 16293: contig of 1254 bp in length

* 16393: gap of unknown length

* 17821: contig of 1428 bp in length

* 17921: gap of unknown length

* 19117: contig of 1196 bp in length

* 19217: gap of unknown length

* 20290: contig of 1073 bp in length

* 20390: gap of unknown length

* 22378: contig of 1988 bp in length

* 22478: gap of unknown length

* 24584: contig of 2106 bp in length

* 24684: gap of unknown length

* 26778: contig of 2094 bp in length

* 26878: gap of unknown length

* 28037: contig of 1159 bp in length

* 28137: gap of unknown length

* 29740: contig of 1603 bp in length

* 29840: gap of unknown length

* 31993: contig of 2153 bp in length

* 32093: gap of unknown length

* 34245: contig of 2152 bp in length

* 34345: gap of unknown length

* 35405: contig of 1060 bp in length

* 35505: gap of unknown length

* 37610: contig of 2105 bp in length

* 37710: gap of unknown length

* 38724: contig of 1014 bp in length

* 38824: gap of unknown length

* 40388: contig of 1564 bp in length

* 40488: gap of unknown length

* 42253: contig of 1765 bp in length

* 42353: gap of unknown length

```

* 42554 45557: contig of 3204 bp in length
* 45558 45657: gap of unknown length
* 45658 49442: contig of 3775 bp in length
* 49433 49532: gap of unknown length
* 49533 51958: contig of 2426 bp in length
* 51959 52058: gap of unknown length
* 52059 54468: contig of 2410 bp in length
* 54469 54569: gap of unknown length
* 54569 57442: contig of 2874 bp in length
* 57443 57542: gap of unknown length
* 57543 60503: contig of 2961 bp in length
* 60504 60603: gap of unknown length
* 60603 63220: contig of 2617 bp in length
* 63221 63320: gap of unknown length
* 63321 66198: contig of 2878 bp in length
* 66199 66298: gap of unknown length
* 66299 68570: contig of 2272 bp in length
* 68571 68671: gap of unknown length
* 68671 72180: contig of 3510 bp in length
* 72181 72280: gap of unknown length
* 72281 75540: contig of 3260 bp in length
* 75541 75640: gap of unknown length
* 75641 79577: contig of 3937 bp in length
* 79578 79677: gap of unknown length
* 79678 82755: contig of 3078 bp in length
* 82756 82855: gap of unknown length
* 82856 85669: contig of 2814 bp in length
* 85670 85769: gap of unknown length
* 85770 89449: contig of 3680 bp in length
* 89450 89549: gap of unknown length
* 89550 92645: contig of 3097 bp in length
* 92646 92745: gap of unknown length
* 92746 96376: contig of 3630 bp in length
* 96377 96476: gap of unknown length
* 96477 99555: contig of 3489 bp in length
* 99556 100055: gap of unknown length
* 100056 104630: contig of 4541 bp in length
* 104631 104706: gap of unknown length
* 104707 110232: contig of 5576 bp in length
* 110233 110283: gap of unknown length
* 110284 112931: contig of 2599 bp in length
* 112932 113031: gap of unknown length
* 113032 117078: contig of 3997 bp in length
* 117079 117178: gap of unknown length
* 117179 122292: contig of 5114 bp in length
* 122293 122392: gap of unknown length
* 122393 125710: contig of 3318 bp in length
* 125711 125810: gap of unknown length
* 125811 132052: contig of 6242 bp in length
* 132053 132152: gap of unknown length
* 132153 136836: contig of 4684 bp in length
* 136837 136937: gap of unknown length
* 136938 142183: contig of 5247 bp in length

```

Query Match 1.1%; Score 21; DB 2; Length 168632;
 Best Local Similarity 100.0%; Pred. No. 10;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1034 ACATTGGATCTCGAAG 1054
 Db 114524 ACATTGGATCTCGAAG 114544

RESULT 26
 LOCUS HSA010770 170008 bp DNA linear PRI 07-JAN-1999
 DEFINITION Homo sapiens hyperion gene, exons 1-50.
 ACCESSION AJ010770
 VERSION AJ010770.1 GI:4128007
 KEYWORDS hyperion gene.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE
 1 (bases 1 to 170008)
 AUTHORS Kemmer, W. A.
 TITLE Direct Submission
 JOURNAL Submitted (31-AUG-1998) Kemmer W.A., Department of Biochemistry,
 Max-Planck Institute for Developmental Biology, Spemannstr. 35/11,
 72076 Tuebingen, GERMANY
 REFERENCE
 2 (bases 1 to 170008)
 AUTHORS Kemmer, W. A., Deiss, S. and Schwarz, U.
 TITLE Cloning of Hyperion
 JOURNAL Unpublished
 FEATURES
 source
 1..170008
 location/Qualifiers
 1..170008
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="7"
 /map="q21-q22"
 511..169570
 /gene="hyperion"
 join(511..558,29884..29919,33122..33379,39700..39744,
 51569..51622,52296..52466,54032..54187,55014..55211,
 60259..62646,71840..72053,73660..73739,75540..75678,
 76428..76513,81649..81763,82225..82420,89306..89402,
 90923..91015,97830..98183,100085..100309,101457..101597,
 102079..102182,104419..104624,112137..112369,
 120671..120833,121685..121897,124642..124874,
 125838..125957,129441..129617,130316..130420,
 132626..136418,137107..137286,138490..139563,
 141933..142073,142581..143106,144230..144351,
 144906..145097,145639..145827,148796..148940,
 152508..152727,154434..154584,156100..156768,
 156997..157205,157520..157625,159098..159280,
 160267..160467,162005..162237,165089..165174,
 166704..166833,167905..168044,169533..169570)
 /gene="hyperion"
 /evidence=experimental
 join(511..558,29884..29919,33122..33379,39700..39744,
 51569..51622,52296..52466,54032..54187,55014..55211,
 60259..62646,71840..72053,73660..73739,75540..75678,
 76428..76513,81649..81763,82225..82420,89306..89402,
 90923..91015,97830..98183,100085..100309,101457..101597,
 102079..102182,104419..104624,112137..112369,
 120671..120833,121685..121897,124642..124874,
 125838..125957,129441..129617,130316..130420,
 132626..136418,137107..137286,138490..139563,
 141933..142073,142581..143106,144230..144351,
 144906..145097,145639..145827,148796..148940,
 152508..152727,154434..154584,156100..156768,
 156997..157205,157520..157625,159098..159280,
 160267..160467,162005..162237,165089..165174,
 166704..166833,167905..168044,169533..169570)
 /gene="hyperion"
 /codon_start=1
 /evidence=experimental
 /product="Hyperion protein"
 /protein_id="CAA09361.1"
 /db_xref="GI:4128008"
 /translation="MEDERQKLENGAKKIEELSLAFVROLAQRKASDQSP
 SKOKKRRKTSKSHVSAHDLNIDOSCENMYNSQSRVSTVPESTIKRTLSG
 EITSHQGFVLESEISTTADCSSEVGCSPVWRTGPGTLLREBEVGVDSYSEQ
 GAQSPFHLHMESELAGKQHEIEELNRELEMRVYGTGEGTQLEFPALIKRQGI
 TQTLRNLQOARREKDETFMRFLTELTEOSQKQIOFOLOASBETLRNSHSTADLL
 OKOOLITHQOOLEBDHLEEDYQKKKEPTQOISTLOKIKITYEBOQKXKNSKE
 EIOERTTIEELNTKIEBEKTLBKDLTTADKLEGLQIOVQKQETKMKLEL
 TNSKQERSSEIRKQMGVLEIRKRNKNSQFETDIYQRMQETRLQELRFLD
 ENYGOIYQMKOLIRQMAQVEMKTRKSGEMENLRQYSINIVNEDIKMNVAIN
 EINKLQDINSQKELKEBGLIEBKCLQGLQLELVELEFSRQIOARQITLAEQ
 ESKLNHAKSLSTVDILKAEIVASRSREKLKHAETVANKIKLEMEKKNAYLD
 RNASESEALERLTQQLSHHEBSKLKEDIEIRRNIEKTKNLTGHVQOQDGL
 QNMSQKITEWQPEKQNLITKQNLILLEISKDLQOQSLVNSKSEMTLQINELQKFI
 EILROEKKEKTEQVQLOLQKTLLEKQMKENDLEKRAQLAENSILKDEKTI
 LEDMLKIHVPVQEEERLIFLDSIKSKSDQSWKEKEIIEERNEEDKQOQILNEIE
 KQNTPSFAKKNFVNVYQLOEYACILKVKDLEDSKQKQLEYYSKILALNEIHL

ORINPTTVKMKSVFDEKTFVAETLMGEVVEKDTTTELMKLELVTKREKLELSQRLS
LDLSQKQKHGEIIFLNEEVKSLQKEKQVSLRCRELEIINHNRRAENVQCDTQVSS
LLDGVVMTSGAGSVSKVNSFGESKIMVEDKVSFENMTVGEESKQELIIDHLP
VNTSESSLRATQPEKNDLQKELNLEQNDRLQMAQRIICLSLTVSTHVDQVREY
MENEKDKALCSKEBELPAQOEKIKELQIHOLELQMTQETGDEGKPLHLLGKQL
KAVSECSYFQTLCSVLGEYVTPALKCEVNAEDKENSVDYISENEDEPELODYVEVQ
DFOENMHTLANKYEVNKLAVLOTRLSKWQOQDGMKLEFGEENLPKETEFLSIH
SOMTNLEIDVNHKSKSSLOQLEKTKLEEQVQELSLISSQQLKETEYQBAEIH
CLNRLQAVSSSTVPPSLPVDVSVITESDAQRTWPGSKKNIDGTIEFSGEFGVKE
ETNIVKLEKQYQOLEEVAKVIVSMAISAFQAQTELSRIISGKENTASSQAQHAVQ
QOEQYFNEMKLSQDQIGQTFQETVDVFKPEEFKPLSKELGHEGKEILLNSNDPHDPE
SKDCVLTISEMFSKQDTFIVQSIHDEIISVSDMSASQMLNKEOLEDMFQELVROY
QEOHQATELLQAQHWOMEROREQEOLOEIKELNQLAQRSSIDNENLVSERVL
LEELEALQSLAGREKLCCLRSSTQNGNENQGEVEQTFKEKELDKRPPDVP
BILSNRVALQKANNRLKILLEVVTAAVEETIGRHVLGILDRSSKQSQSSALWR
SEASVKSCHVEHTRVTSIPSYSGDMRNDINMWSKVTEGETELSRULRVQF
AGTEIDPENEELMNIISRLQAQAEKLEAISSETSSOLEHAKVTQTELMRESFPQKE
ATESLKQOEELRLHESRAREOLAVELSKAEGVIDGVADEKTLFERQIOEKTDIID
RLEOBLLCASNRLOLEAEQOIOEERELLSROKEMKAAGPVQEQLLQTEKLMKE
KLEVQCAQKVRDLQKVKALEIDVEQVSRFIELEQKNTLMDLRQQAQLEKQL
EKMRGLDEQAIIDREHERDVFQETIQLEQKVVPVPOPISEHOTREVEOLANHLKE
KTDKSELLSKEQLOQDIQERNEIEKLEFRVRELEQALLVSADTFQKVDNRHFGA
VEAKPELSLEVQOAEARDAIDRKEKEITNLSEOLEQEPRELEKNNEEVQOOLHOLEIQ
KKESTTRLOELEQENKLPKDDMEKLGIAIKESDAMSTODOHVLFGKFAQIIOEKEVI
DQLEAOTVKLQOQKITTDNKKVIREKHELIRDLETQIECLMSDQECVKRREBIEQL
NEVIEKQOELANTQKTSMAHLSSEADSLKHQLOVVIKALALSQQVETANEEMT
FMKVLKETFNMQLTOELFSLKRERESVEKIQSIPENSVNVAIDHLSKDKPKEVY
LTEDALKSLENQTFYKFPENGKSIINLETRLQLSESTVSAKLELTQCYQIKDMQ
EQQOFETEMLOKQTVNLOKIVVEKVAALVSOIOLEAVQYAKFCQDNQOTISSPERT
NIONLNRLDELSGDISALTRLISELESQVEMHTSLILEKEQVEIAEKNVLEKX
LLELOKLLGEGKQREKREKRSQDVVKTTELPHSNEESGFFNEALAESVA
TKALASVKEAKQLQELLVKTENMTSLQKDLQVDRHLAAEKULSILEKDETEV
QESKACMFPLPKLSKSIASQDGTIKTSSNQTPQILVNAGIQINLOQSCSSE
VTIISQFTEKIQMQLHAAEILDMESRHSITETTLKREHYAVQLLKESCGTKAV
IQCLRSKEVFGYNNCFSTLCDSGSDMOGIVLTHSQCFDIASBGRGEESASDSEF
KKIKGLLRVAVNEGMVLSLTPESVSDGEDHSIOQVSEPMLEERKAVINTTSSIKDIL
TKMOLQREAVYVDSQSHESFDMRGELLALLQVFLREERSVLLAATFELTALGTID
AVGLNCLLEQRIQSGVEYQAMCELCQADRRSLSEIOALHAQNMGRKITYLKREQS
QSGOELLLEYNIQKQSQMLEQVLSMDRATLQELQSSKQVVAELKSELAQTK
LELETTLKAQHKHLEAFRLVDKTDVHLLNDTLASQKRSRELQWALEKEKAK
LGRSEERKDELEDKFSLEQKQKQNLNLLLEQKQNLNLEQKQKTESQORMLYDAOL
SEQORNLLOVLESEKVRIREMSSITLDREREHLAQLQSSDGTGQSRPPLPSRDLK
ELQOLKEKHSRIVELLNETEKYKLDLSLQTRQMEKQBRVHRTKLQTEQANTBQKK
MHELQKVEDLQRLQEKQVQYKLDLEQQLQIMQEFQKQELEREKRESRLIYQ
NLNEFTWLSLSDTRNNVLQKIEGTESKESYAKLIEMNGGGTGCHNEHEMLRQKQ
CVASQLVLPKASERIQFETADDEFTWQENIDEIILQKLTGOQGEPSLVSPS
TSCGSLTERLLRQNAELTGHISQLTEKNDLRNMVKLEQIRWYROTGAGRDNSSRF
SLNGANIEALIAEKEVWREKTLQKSLKBAEVYKLAELRNLSLLOTLSPDSE
HVTLKRIYKYLRAESFRKALIYKKYLLLLLGFQCEADATALLARMGQGPATDL
EVTNRPKGFTRFSARVRSIAIRMKPLVRMRHRTVGSVSNINLRMGFGLNQAEKT
DSFYHSSGGGLYEPRHTYRSRDLDIRPLPQNRYPGTADFPNPGSLACSQLQ
NYDPDRALTDYITRLEALQRLGTTQSGSTTQFHAGMRR"
511..558
/gene="hyperion"
/number=1
559..29883
/gene="hyperion"
/number=1
29894..29919
/gene="hyperion"
/number=2
29920..33121
/gene="hyperion"
/number=2
33122..33379
/gene="hyperion"
/number=3
33380..39699
/gene="hyperion"
/number=3
39700..39744
/gene="hyperion"
/number=4

intron 39745..51568
/gene="hyperion"
/number=4
exon 51569..51622
/gene="hyperion"
/number=5
intron 51623..52295
/gene="hyperion"
/number=5
exon 52296..52466
/gene="hyperion"
/number=6
intron 52467..54031
/gene="hyperion"
/number=6
exon 54032..54187
/gene="hyperion"
/number=7
intron 54188..55013
/gene="hyperion"
/number=7
exon 55014..55211
/gene="hyperion"
/number=8
intron 55212..60258
/gene="hyperion"
/number=8
misc_feature 60112..60258
/gene="hyperion"
/note="isoform 1-5A"
exon 60259..62646
/gene="hyperion"
/number=9
misc_feature 60525..61129
/gene="hyperion"
/note="similarity to M.musculus EST C79026 (NID:92519356)"
60713..61137
/gene="hyperion"
/note="similarity to M.musculus EST AAS42309 (NID:92288743) vk53c08.r1"
60736..61049
/gene="hyperion"
/note="match to human EST AA044975 (NID:91523178) zf53g11.r1"
61106..61536
/gene="hyperion"
Query Match 1.1%; Score 21; DB 9; Length 170008;
Best Local Similarity 100.0%; Pred.No.10;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1395 ATTCTTTTTCGAGAAATTC 1415
|||||
Db 95395 ATTCTTTTTCGAGAAATTC 95415
|||||
RESULT 27
AC096466/c
LOCUS
DEFINITION Rattus norvegicus clone CH230-84G6, *** SEQUENCING IN PROGRESS ***,
60 unordered pieces.
ACCESSION AC096466
VERSION AC096466.3 GI:21723957
KEYWORDS HTG; HTGS PHASE1.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 175809)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Aisbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,

Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P.,
 Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C.,
 Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,
 Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,
 Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,
 Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,
 Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H.,
 Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,
 Earhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,
 Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,
 Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,
 Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,
 Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J.,
 Hernandez, O., Hodgson, A., Hogue, M., Holloway, C., Hollins, B.,
 Homsi, F., Howard, S., Huber, U., Hui, S., Hume, J., Jackson, L.E.,
 Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudan, S.,
 Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,
 Kratoch, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,
 Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Louised, H.,
 Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,
 Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E.,
 Massey, E., Mawhinney, E., McLeod, M.P., Meador, M., Mei, G., Metzger, M.,
 Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,
 Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, N.,
 Nguyen, N., Nickerson, E., Nwokenkwo, S., Ogun, M., Okunolu, G.,
 Oragunye, N., Oyiedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,
 Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y.,
 Rivers, M., Rojas, A., Rojudo, I., Rolfe, M., Ruiz, S., Savery, G.,
 Scherer, S., Scott, G., Shen, H., Shoochari, N., Sisson, I.,
 Sodergren, E., Sonike, T., Sparks, A., Stanley, H., Stone, H.,
 Sutton, A., Svarek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,
 Tansley, D., Taylor, C., Taylor, T., Telford, B., Thomas, R., Thomas, S.,
 Usmami, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wang, Q.,
 Wang, S., Ward-Moore, S., Warren, R., Washington, C., Wallington, S.,
 Williams, G., Williamson, A., Wleczky, R., Woodson, S., Worley, K.,
 Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,
 Weinstein, G., and Gibbs, R.

Direct Submission
 Unpublished
 2 (bases 1 to 175809)
 Worley, K.C.

Direct Submission
 Submitted (17-SEP-2001) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 175809)

Direct Submission
 Submitted (11-JUL-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Jul 10, 2002 this sequence version replaced gi:17956421.

----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu

----- Project Information
 Center project name: GFCO
 Center clone name: CH230-84G6

----- Summary Statistics
 Sequencing vector: Plasmid
 Chemistry: Dye-terminator Big Dye: 100% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 122727 bases at least Q40
 Consensus quality: 126253 bases at least Q30
 Consensus quality: 129204 bases at least Q20

 * NOTE: Estimated insert size may differ from sequence length
 * (see <http://www.hgsc.bcm.tmc.edu/docs/genbankdraftdata.html>).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 60 contigs. The true order of the pieces
 * is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1
 1082: contig of 1082 bp in length
 1182: gap of unknown length
 2774: contig of 1592 bp in length
 2874: gap of unknown length
 4543: contig of 1669 bp in length
 4544: gap of unknown length
 5762: contig of 119 bp in length
 5862: gap of unknown length
 5863: contig of 1076 bp in length
 6939: gap of unknown length
 7039: contig of 1417 bp in length
 8456: gap of unknown length
 8556: contig of 1136 bp in length
 9691: gap of unknown length
 9791: gap of unknown length
 9992: contig of 1176 bp in length
 10967: gap of unknown length
 11067: gap of unknown length
 12717: contig of 1650 bp in length
 12718: gap of unknown length
 14511: contig of 1734 bp in length
 14552: gap of unknown length
 14552: contig of 1394 bp in length
 16045: gap of unknown length
 16145: gap of unknown length
 16146: contig of 1539 bp in length
 17684: gap of unknown length
 17784: contig of 1100 bp in length
 18884: gap of unknown length
 18984: gap of unknown length
 20769: contig of 1785 bp in length
 20869: gap of unknown length
 22794: contig of 1925 bp in length
 22795: gap of unknown length
 22895: contig of 1114 bp in length
 24008: gap of unknown length
 24109: contig of 1445 bp in length
 25593: gap of unknown length
 25594: gap of unknown length
 27286: contig of 1593 bp in length
 27287: gap of unknown length
 27387: contig of 1369 bp in length
 28756: gap of unknown length
 28855: gap of unknown length
 30486: contig of 1631 bp in length
 30586: gap of unknown length
 32208: contig of 1622 bp in length
 32308: gap of unknown length
 32309: gap of unknown length
 33636: contig of 1328 bp in length
 33637: gap of unknown length
 33737: contig of 1232 bp in length
 34968: gap of unknown length
 35068: gap of unknown length
 35069: contig of 1836 bp in length
 37004: gap of unknown length
 39237: contig of 2233 bp in length
 39337: gap of unknown length
 39338: gap of unknown length
 40459: contig of 1122 bp in length
 40559: gap of unknown length
 42802: contig of 2143 bp in length
 42803: gap of unknown length
 42803: gap of unknown length
 46164: contig of 3362 bp in length
 46264: gap of unknown length
 48471: contig of 2207 bp in length
 48472: gap of unknown length
 51355: contig of 2784 bp in length
 51356: gap of unknown length
 51455: gap of unknown length
 53805: contig of 2350 bp in length
 53905: gap of unknown length
 53906: gap of unknown length
 55202: contig of 1237 bp in length
 55203: gap of unknown length
 57937: contig of 2635 bp in length
 58037: gap of unknown length
 61502: contig of 3465 bp in length
 61502: gap of unknown length

```
* 61603 63399: contig of 1797 bp in length
* 63400 63499: gap of unknown length
* 63500 63599: contig of 3008 bp in length
* 66508 66607: gap of unknown length
* 66608 66607: contig of 3023 bp in length
* 69631 69730: gap of unknown length
* 71599 71599: contig of 1869 bp in length
* 71600 71699: gap of unknown length
* 71700 71729: contig of 2430 bp in length
* 74130 74229: contig of 2415 bp in length
* 74230 76644: contig of 2415 bp in length
* 76645 76744: gap of unknown length
* 76745 79598: contig of 2854 bp in length
* 79599 79698: gap of unknown length
* 79699 84161: contig of 4463 bp in length
* 84162 84261: gap of unknown length
* 84262 87349: contig of 3088 bp in length
* 87350 87449: gap of unknown length
* 87450 91094: contig of 3645 bp in length
* 91095 91194: gap of unknown length
* 91195 95118: contig of 3924 bp in length
* 95119 95218: gap of unknown length
* 95219 100153: contig of 4935 bp in length
* 100154 100253: gap of unknown length
* 100254 103585: contig of 3332 bp in length
* 103586 103685: gap of unknown length
* 103686 109212: contig of 5527 bp in length
* 109213 109312: gap of unknown length
* 109313 113568: contig of 4256 bp in length
* 113569 113668: gap of unknown length
* 113669 118480: contig of 4812 bp in length
* 118481 118580: gap of unknown length
* 118581 124124: contig of 5544 bp in length
* 124125 124224: gap of unknown length
* 124225 129675: contig of 5451 bp in length
* 129676 129775: gap of unknown length
* 129776 132662: contig of 5487 bp in length
* 132663 135362: gap of unknown length
* 135363 140677: contig of 5315 bp in length
```

```
Query Match 1.1% Score 21; DB 2; Length 175809;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 1643 GTCCTTAATTTGCAATTATT 1663
|||||
Db 86185 GTCCTTAATTTGCAATTATT 86165
```

```
RESULT 28
BSUB0011/c
LOCUS
DEFINITION
  BSUB0011 207730 bp DNA linear BCT 26-NOV-1997
  Bacillus subtilis complete genome (section 11 of 21): from 2000171
  to 207900.
Z99114 AL009126
Z99114.1 GI:2634230
KEYWORDS
SOURCE
  Bacillus subtilis.
ORGANISM
  Bacillus subtilis
  Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
REFERENCE
  1 (bases 1 to 207730)
  Kunst, F., Ogasawara, N., Moszer, I., Albertini, A.M., Alloni, G.,
  Azevedo, V., Bertero, M.G., Bessieres, P., Bolotin, A., Borchert, S.,
  Borriss, R., Boursier, L., Brans, A., Braun, M., Brignell, S.C.,
  Bron, S., Brouillet, S., Bruch, C.V., Caldwell, B., Capuano, V.,
  Carter, N.M., Choi, S.K., Codani, J.J., Conner, I.F., Cummings, N.J.,
  Daniel, R.A., Denzot, F., Devine, K.M., Dusterhoft, A., Ehrlich, S.D.,
  Emerson, P.T., Entian, K.D., Errington, J., Fabret, C., Ferrari, E.,
  Foulger, D., Fritz, C., Fujita, M., Fujita, Y., Fuma, S., Galizzi, A.,
  Galleron, N., Ghim, S.Y., Glaser, P., Goffeau, A., Golligly, E.J.,
  Grandi, G., Guiseppe, G., Guy, B.J., Haga, K., Haele, J., Harwood, C.R.,
  Henaut, A., Hilbert, H., Holsappel, S., Hosono, S., Hullo, M.F.,
  Itaya, M., Jones, L., Joris, B., Karamata, D., Kasehara, Y.,
```

```
Klaerr-Blanchard, M., Klein, C., Kobayashi, Y., Koetter, P.,
Koningstein, G., Krogh, S., Kumano, M., Kurita, K., Lapidus, A.,
Lardinois, S., Lauber, J., Lazarevic, V., Lee, S.M., Levine, A., Liu, H.,
Masuda, S., Maue, C., Medigue, C., Medina, N., Mellado, R.P.,
Mizuno, M., Moestl, D., Nakai, S., Noback, M., Noone, D., O'Reilly, M.,
Ogawa, K., Ogiwara, A., Oudega, B., Park, S.H., Paro, V., Pohl, T.M.,
Portetle, D., Porwollik, S., Prescott, A.M., Prescan, E., Pujic, P.,
Purnelle, B., Rapoport, G., Rey, M., Reynolds, S., Rieger, M.,
Rivolta, C., Rocha, E., Roche, B., Rose, M., Sadale, Y., Sato, T.,
Savolainen, C., Schleich, S., Schroeter, R., Scoffone, P., Sekiguchi, J.,
Sekowska, A., Seror, S.J., Seror, P., Shin, B.S., Soldo, B.,
Sorokin, A., Tacconi, E., Takagi, T., Takahashi, H., Takemaru, K.,
Takeuchi, M., Tamakoshi, A., Tanaka, T., Terpstra, P., Tognoni, A.,
Tosato, V., Uchiyama, S., Vandenbol, M., Vannier, F., Vassarotti, A.,
Viari, A., Wambutt, R., Wedler, E., Wedler, H., Weitzensegger, T.,
Winters, P., Wipat, A., Yamamoto, H., Yamane, K., Yamamoto, K., Yata, K.,
Yoshida, K., Yoshikawa, H.F., Zumstein, E., Yoshikawa, H. and
Danchin, A.
The complete genome sequence of the gram-positive bacterium
Bacillus subtilis
Nature 390 (6657), 249-256 (1997)
98044033
PUBMED
9384377
REFERENCE
2 (bases 1 to 207730)
Kunst, F., Ogasawara, N., Yoshikawa, H. and Danchin, A.
Direct Submission
Submitted (18-NOV-1997) I. Moszer, A. Danchin, Institut Pasteur,
Regulation de l'Expression Genetique, 28 rue du Docteur Roux, 75724
Paris Cedex 15, FRANCE. E-mail: moszer@pasteur.fr,
adanchin@pasteur.fr Phone: +33 (0)1 45 68 84 41, Fax: +33 (0)1 45
68 89 48
FEATURES
source
  1. 207730
  /organism="Bacillus subtilis"
  /strain="168"
  /db_xref="taxon:1423"
  complement (10. 1436)
  /genes="yoeA"
  complement (10. 33)
  /genes="yoeA"
  complement (45. 1436)
  /genes="yoeA"
  /function="unknown"
  /notes="similar to hypothetical proteins from B. subtilis"
  /codon_start=1
  /transl_table=11
  /protein_id="CAB13730.1"
  /db_xref="GI:2634231"
  /db_xref="SPTRMBL:O34474"
  /translation="METSQEPVLTQKSKPSVWMSLFLVPLLLSNVLSQGVGM
  MAVRWLGVDAVAASVFFPFLFLLISFTIGIGSSILIGQAVGAKNEERLKVCT
  TLTFLLGLVLAIGSIFLTDILRLMCTPENVIHVSANVARTLVAMPFPLVAVT
  TFLRGDSKTPYTLIVSTVINIALPVLVLLMFGPKLGYGSAVATVISTATFL
  VLMVLRKRRHLPQDFDKRRYLRKMDKELLVLLRLGVSPASINLVLSLSEIAVISFV
  NYIIGNATAYGVNQVAVSQVPAVSLGIAVSIFAAQSIGANEFRLKQVIRATNR
  NYIIGNATAYGVNQVAVSQVPAVSLGIAVSIFAAQSIGANEFRLKQVIRATNR
  ASGTVLWPTVISIFAIWGVPEVPAFVLVSHVTKLEILGVWVGYPAFAVSLLVGYVQ
  FVKKKQITRLIQ"
  1722. 2345
  /genes="yoeB"
  1722. 2267
  /genes="yoeB"
  /function="unknown"
  /codon_start=1
  /transl_table=11
  /protein_id="CAB13731.1"
  /db_xref="GI:2634232"
  /db_xref="SPTRMBL:O34841"
  /translation="MKKCLLFLTTIALILSLSTNAFAKNTSGDLSQKQALQALSARE
  HFNTWSQNPQKVKAVCVSGTPEYQNVYVMSDLCTKAKAVNLTPLPTKATBK
  GFNDYHTVSKGLAVPIGDGNLLNKKSTAKLISKGGSTIITEFTVPLDGGSPSAK
  RKVTYFKENKKVKNQFDVVI"
  2274. 2292
  terminator
```

```

/gene="yosB"
319. .2315
/terminator
/gene="yosB"
2359. .2434
/crna
/gene="trnSL-Arg1"
2359. .2434
/product="trnSL-Arg1"
/gene="trnSL-Arg1"
complement(2745. .3032)
/gene="yocC"
complement(2745. .3032)
/gene="yocC"
/function="unknown"
/codon_start=1
/transl_table=11
/protein_id="CAB13732.1"
/db_xref="GI:2634233"
/db_xref="SPTREMBL:O35009"
/translacion="MHIVQPIRELEKIOEVKOYLANKKRDYFLPIFGINSALRISDI
LPIQVADVCKNKHMLNATESONEKEKDSIFRLAETGNINIVERHEBKRI5F"
complement(3327. .3578)
/gene="yocD"
complement(3327. .3578)
/terminator
/gene="yocD"
complement(3348. .3578)
/gene="yocD"
/function="unknown"
/codon_start=1
/transl_table=11
/protein_id="CAB13733.1"
/db_xref="GI:2634234"
/db_xref="SPTREMBL:O34555"
/translacion="MYTIEIETAEITNLSDEYIKSLVBEKIRAVHDGEQFLYKQEP
KTHLEQENYKALVOETLINEPIEDIDVDED"
3763. .5526
/gene="yocE"
3763. .5526
/gene="yocE"
/EC_number="2.3.2.2"
/function="glutathione metabolism"
/note="alternate gene name: pac"
/codon_start=1
/transl_table=11
/product="gamma-glutamyltranspeptidase"
/protein_id="CAB13734.1"
/db_xref="GI:2634235"
/db_xref="SWISS-PROT:P54422"
/translacion="MKRTMVCULFALSVLIVAGSVFPFAEAKKPKSYDEYKQVDVG
KDCMVAIARHPLASRIGDVKKGNALDAVALIOPLNTEPMMSGIGGGFMVYDQ
KITDITLIDSRERAPACATPDMFLDENGKAIIPSEBVTGTAIGVPTGLBEALDK
WGRSMKQLITPSTIKLAKGFPIDSVLAELISDOEKLSTAKDVFPLNGEPLKGD
TLIQDLAKTFKILRSKGTDAFYKGFAKTISDVQDFGSMTEKLENDITIDEPI
WGDYGOIATTPPSSGGIFLLQMLKILHFNLSQYDVRSEWERYQLAETMHLVAD
RASYAGDPEFNVPLKGLLHFDYIKERQILINDQVKKPKAGDPMKYQSGSNYKQV
EOPKDKVESCOTHTTAVDWRGNVSVYTTIEOLGEGINVPDYGVLINLLETPEALP
GGANEVQPKRPLUSSMTPTILFKDPRVILVVGSGAGTISSVLOTLIHYIEWELK
AAVEEPIYTNSSSYREDEGVPDVLKLMGMRKGTSPVDIGNVQISIDHENCT
FKGVADSSRNGAIGINLKRK"
complement(5535. .6484)
/gene="yofA"
complement(5535. .6484)
/gene="yofA"
complement(5627. .6484)
/gene="yofA"
/function="unknown"
/note="similar to transcriptional regulator (LysR family)"
/codon_start=1
/transl_table=11
/protein_id="CAB13735.1"
/db_xref="GI:2634236"
/db_xref="SPTREMBL:O34685"
/translacion="WESGDLKIFQAVAREGSIYRAQOMLVQSVNVARVNLLEDIN
IRLFRTRMGKLTAAENLQIYADQVSLIDQAEKSTRMSRQKXGPLRIGSLTMAV

```

```

THLPEHAASFLRRPEVDSLVTADTHHLIQVLIDHKVDCAFVYGPVEHAAROLHVS
HDEILVLSREGTAEIDMIQOPMLFFGAGCSHRDKLLEBAGIHNKITEFGLKAI
IKGVASAGMTALLPKSAVDGSEHRTNWIMHOLPPSYODLEIVFIYRDFITSAPQTF
LDEINEMKR"
6612. .7601
/gene="yogA"
6612. .7601
/gene="yogA"
/function="unknown"
/note="similar to alcohol dehydrogenase"
/codon_start=1
/transl_table=11
/protein_id="CAB13736.1"
/db_xref="GI:2634237"
/db_xref="SPTREMBL:O35017"
/translacion="MKAVIHNGKAGLGLSVQDVPSTKPGYGEVVKLKSGALNHRDL
FLMKRSQDQDHMLIGSDGAGITIEIGGVKNVWYQFVVIPEPLNMDLTENVVPE
ILGGSDDTLAEVYITPSQNAIKKPAIYSHBEAGVLPISALTARALPTKOLKQGH
LLIPIGSAVATYALFMKALIGATVSVTSRSEERKALKGADYAFDPSYNMBOLO
GKIDIVLDSIGPALFSEYFPHVKNRGIYVFGASGDNLSFPVRSLEFPVNVLTGS
MGSGEFOAMLAFLDKHLRPVIDRIYFLERACEAYKKMOGRORGNIGIYME"
complement(7611. .9139)
/gene="glbB"
complement(7611. .7625)
/gene="glbB"
complement(7658. .9139)
/gene="glbB"
/EC_number="1.4.1.13"
/function="glutamate biosynthesis"
/codon_start=1
/transl_table=11
/product="glutamate synthase (small subunit)"

```

```

Query Match 1.1% Score 21; DB 1; Length 207730;
Best Local Similarity 100.0% Pred. No. 10;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 749 CAACATACACAGTTCATATG 769
Db 115484 CAACATACACAGTTCATATG 115464

```

```

RESULT 29
AL732291 207890 bp DNA linear HTG 17-AUG-2002
LOCUS AL732291
DEFINITION Mus musculus chromosome X clone RP23-48114, *** SEQUENCING IN
PROGRESS ***, 2 unordered pieces.
ACCESSION AL732291
VERSION AL732291.6 GI:21615699
KEYWORDS HTG; HTGS PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE house mouse.
ORGANISM Mus musculus.
REFERENCE 1 (bases 1 to 207890)
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE Leonagomollet, D.
JOURNAL Direct Submition
Submitted (16-AUG-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Jun 26, 2002 this sequence version replaced gi:2161687.
COMMENT ----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
Project Information
Center project name: BM48114
----- Summary Statistics
Assembly program: XGAP4, version 4.5
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator; 2% of reads
Chemistry: Dye-terminator Big Dye; 97% of reads

```



```

Consensus quality: 207379 bases at least Q40
Consensus quality: 207605 bases at least Q30
Consensus quality: 207722 bases at least Q20
Insert size: 207790; sum-of-contigs
Quality coverage: 9.87x in Q20 bases; sum-of-contigs Quality
coverage: 10.07x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 9063: contig of 9063 bp in length
*
* 9064 9163: Gap of 100 bp
*
* 9164 207890: contig of 198727 bp in length.
*
* Location/Qualifiers
*   1..207890
*     /organism="Mus musculus"
*     /db_xref="taxon:10090"
*     /chromosome="X"
*     /clone="RP23-48I14"
*     /clone_lib="RPC1-23"
*   1..9063
*     /note="assembly fragment:01288"
*     /fragment_chain="1"
*   9164..207890
*     /note="assembly fragment:04457"
*     /fragment_chain="1"
*
* BASE COUNT 67167 a 38739 c 37935 g 63948 t 101 others
* ORIGIN
*
* Query Match
* Best Local Similarity 100.0%; Pred. No. 10;
* Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
*
* Oy 530 CAAAGTCTTAAAAATCTGTA 550
* |||||
* Db 119212 CAAAGTCTTAAAAATCTGTA 119332
* |||||
*
* RESULT 30
* AC009630/c
* LOCUS
* DEFINITION Homo sapiens chromosome 8, clone RP11-360L9, complete sequence.
* ACCESSION AC009630
* VERSION AC009630.14 GI:19848415
* KEYWORDS HTG.
* SOURCE
* ORGANISM Homo sapiens
* Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
* Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
*
* 1 (bases 1 to 219622)
* Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
* Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A.,
* Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,
* Cooke,P., Dearellano,K., Depayre,E., Devon,K., Dewar,K.,
* Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C.,
* Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,
* Hagos,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,
* Karatas,A., Lehoczy,J., Lieu,C., Locke,K., Macdonald,P.,
* Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,
* Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
* Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,
* Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,
* Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
* Tefaye,S., Torruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,
* Wheeler,U., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
*
* Direct Submission
* Submitted (28-AUG-1999) Whitehead Institute/MIT Center for Genome
* Research, 320 Charles Street, Cambridge, MA 02141, USA
* 3 (bases 1 to 219622)
* Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
* Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavsky,L.,
* Boukigalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
* Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
* Cooke,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
* Fero,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
* Gande,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
* Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
* Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
* Landers,T., Lehoczy,J., Levine,R., Lindblad-Toh,K., Liu,G.,
* MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
* McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
* Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
* Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
* Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
* Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
* Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schuback,R.,
* Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
* Strauss,N., Subramanian,A., Talamas,J., Tefaye,S., Theodore,J.,
* Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
* Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
* Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
*
* Direct Submission
* Submitted (22-MAR-2002) Whitehead Institute/MIT Center for Genome
* Research, 320 Charles Street, Cambridge, MA 02141, USA
* 4 (bases 1 to 219622)
* Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
* Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavsky,L.,
* Boukigalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
* Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
* Cooke,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
* Fero,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
* Gande,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
* Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
* Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
* Landers,T., Lehoczy,J., Levine,R., Lindblad-Toh,K., Liu,G.,
* MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
* McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
* Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
* Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
* Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
* Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
* Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schuback,R.,
* Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
* Strauss,N., Subramanian,A., Talamas,J., Tefaye,S., Theodore,J.,
* Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
* Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
* Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
*
* Direct Submission
* Submitted (30-MAR-2002) Whitehead Institute/MIT Center for Genome
* Research, 320 Charles Street, Cambridge, MA 02141, USA
* On Mar 30, 2002 this sequence version replaced gi:19683283.
* All repeats were identified using RepeatMasker:
* Smit, A.F.A. & Green, P. (1996-1997)
* http://ftp.genome.washington.edu/RM/RepeatMasker.html
*
* ----- Genome Center
* Center: Whitehead Institute/ MIT Center for Genome Research
* Center code: WIBR
* Web site: http://www-seq.wi.mit.edu
* Contact: sequence_submissions@genome.wi.mit.edu
* ----- Project Information
* Center project name: L2093
* Center clone name: 360_L_9
* -----
* Location/Qualifiers
* 1..219622
* /organism="Homo sapiens"

```

```

/db_xref="taxon:9606"
/chromosome="8"
/map="8"
/clone="RP11-360L9"
/clone_lib="RPC1-11 Human Male BAC"
complement(740..835)
/rpt_family="MIR"
repeat_region
1626..1730
/rpt_family="GA-rich"
repeat_region
1995..2019
/rpt_family="GC-rich"
2306..2368
/rpt_family="GA-rich"
complement(3506..3702)
/rpt_family="L2"
3987..4054
/rpt_family="MIR3"
5149..5504
/rpt_family="AluSx"
5900..6166
/rpt_family="AluSg"
6260..6432
/rpt_family="MIR"
6482..6363
/rpt_family="MLT1A1"
6877..7352
/rpt_family="MIR"
7370..7571
/rpt_family="AluSx"
7817..7914
/rpt_family="MERSA"
8016..8126
/rpt_family="MSTB1"
8117..8182
/rpt_family="MSTB1"
8183..8489
/rpt_family="AluSx"
8490..8555
/rpt_family="MSTB1"
9370..9577
/rpt_family="AluSx"
complement(9719..9832)
/rpt_family="MIR"
9928..10135
/rpt_family="MIR"
10239..10348
/rpt_family="L2"
10751..11044
/rpt_family="AluDb"
complement(11087..11322)
/rpt_family="L1MA7"
11324..11411
/rpt_family="AluY"
11412..11695
/rpt_family="AluSc"
complement(11696..12441)
/rpt_family="L1MA7"
12442..12744
/rpt_family="AluSx"
complement(12745..12803)
/rpt_family="L1MA7"
12804..13115
complement(13116..13499)
/rpt_family="AluSx"
13500..13792
/rpt_family="AluSx"
complement(13793..14061)
/rpt_family="L1MA7"
complement(14093..14239)
/rpt_family="MIR3"
14535..14557
/rpt_family="(GGGGA)n"

```

```

repeat_region complement(14642..14785)
/rpt_family="L1PA8"
repeat_region 15685..15702
/rpt_family="G-rich"
repeat_region complement(16451..16575)
/rpt_family="MER41A"
repeat_region 16576..16871
/rpt_family="AluSx"
complement(16872..17310)
/rpt_family="MER41A"
repeat_region 19693..19988
/rpt_family="AluDb"
repeat_region 19992..20023
/rpt_family="Alu"
repeat_region 20024..20288
/rpt_family="AluSg/x"
20289..20300
/rpt_family="Alu"
21314..22905
/rpt_family="SVA"
24485..24786
/rpt_family="AluSg"
complement(25042..25134)
/rpt_family="MIR3"
complement(26856..26938)
/rpt_family="MIR"
28565..28766

Query Match 1.1%; Score 21; DB 9; Length 219622;
Best Local Similarity 100.0%; Pred.No. 10;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Qy 654 GGTTTGCTTGACATTGATTT 674

Db 145761 GGTTTGCTTGACATTGATTT 145741

```

RESULT 31
LOCUS AL365337 257967 bp DNA linear HTG 19-JUN-2002
DEFINITION Mus musculus chromosome 17 clone CT7-BM104010, *** SEQUENCING IN
PROGRESSES *** 39 unordered pieces.
ACCESSION AL365337
VERSION AL365337.5 GI:14575247
KEYWORDS HTG; HTGS_PHASE1.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 257967)
Sims,S.
Direct Submission
Submitted (17-JUN-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquerry@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk
On Jun 28, 2001 this sequence version replaced gi:9994249.

----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquerry@sanger.ac.uk
----- Project Information
Center project name: BM104010
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; 108752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 234414 bases at least Q40
Consensus quality: 243875 bases at least Q30
Consensus quality: 249519 bases at least Q20
Insert size: 254167; sum-of-contigs
Insert size: 120873; 44.0% error; agarose-fp
Quality coverage: 3.08x in Q20 bases; sum-of-contigs Quality

```

coverage: 8.64x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
* consists of 39 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 2688: contig of 2688 bp in length
* 2689 2788: gap of 100 bp
* 2789 9326: contig of 6538 bp in length
* 9327 9426: gap of 100 bp
* 9427 12321: contig of 2895 bp in length
* 12322 12421: gap of 100 bp
* 12422 15203: contig of 2782 bp in length
* 15204 15303: gap of 100 bp
* 15304 17463: contig of 2160 bp in length
* 17464 17563: gap of 100 bp
* 17564 19968: contig of 2405 bp in length
* 19969 20068: gap of 100 bp
* 20069 24483: contig of 4415 bp in length
* 24484 24583: gap of 100 bp
* 24584 34118: contig of 9535 bp in length
* 34119 34218: gap of 100 bp
* 34219 43750: contig of 9532 bp in length
* 43751 43850: gap of 100 bp
* 43851 46349: contig of 2499 bp in length
* 46350 46449: gap of 100 bp
* 46450 49796: contig of 3347 bp in length
* 49797 49896: gap of 100 bp
* 49897 53042: contig of 3146 bp in length
* 53043 53142: gap of 100 bp
* 53143 56709: contig of 3567 bp in length
* 56710 56809: gap of 100 bp
* 56810 57571: contig of 18942 bp in length
* 57572 75851: gap of 100 bp
* 75852 98190: contig of 22339 bp in length
* 98191 98290: gap of 100 bp
* 98291 100572: contig of 2282 bp in length
* 100573 100672: gap of 100 bp
* 100673 109672: contig of 9000 bp in length
* 109673 109772: gap of 100 bp
* 109773 118329: contig of 8557 bp in length
* 118330 118429: gap of 100 bp
* 118430 123121: contig of 4692 bp in length
* 123122 123221: gap of 100 bp
* 123222 136596: contig of 13375 bp in length
* 136597 136696: gap of 100 bp
* 136697 139702: contig of 3006 bp in length
* 139703 139802: gap of 100 bp
* 139803 148966: contig of 9164 bp in length
* 148967 149066: gap of 100 bp
* 149067 151784: contig of 2718 bp in length
* 151785 151884: gap of 100 bp
* 151885 154359: contig of 2475 bp in length
* 154360 154459: gap of 100 bp
* 154460 156467: contig of 2008 bp in length
* 156468 156567: gap of 100 bp
* 156568 160930: contig of 4363 bp in length
* 160931 161030: gap of 100 bp
* 161031 165186: contig of 4156 bp in length
* 165187 165286: gap of 100 bp
* 165287 168030: contig of 2744 bp in length
* 168031 168130: gap of 100 bp
* 168131 208608: contig of 40478 bp in length
* 208609 208708: gap of 100 bp
* 208709 212266: contig of 3558 bp in length
* 212267 212366: gap of 100 bp
* 212367 221813: contig of 9447 bp in length
* 221814 221913: gap of 100 bp
* 221914 228741: contig of 6828 bp in length

* 228742 228841: gap of 100 bp
* 228842 234071: contig of 5230 bp in length
* 234072 234171: gap of 100 bp
* 234172 236386: contig of 2215 bp in length
* 236387 236486: gap of 100 bp
* 236487 239666: contig of 3180 bp in length
* 239667 239766: gap of 100 bp
* 239767 244113: contig of 4347 bp in length
* 244114 244213: gap of 100 bp
* 244214 247099: contig of 2886 bp in length
* 247100 247199: gap of 100 bp
* 247200 249955: contig of 2756 bp in length
* 249956 250055: gap of 100 bp
* 250056 257967: contig of 7912 bp in length.
FEATURES
 source
 1. .257967
 /organism="Mus musculus"
 /db xref="taxon:10090"
 /chromosome="17"
 /clone="CT7-BM104010"
 /clone_lib="CITB-CJ7-B"
 1. .2688
 /notes="assembly fragment:01624"
 fragment_chain:1
 2789. .9326
 /notes="assembly fragment:03057"
 fragment_chain:1
 9427. .12321
 /notes="assembly fragment:03136"
 fragment_chain:1
 12422. .15203
 /notes="assembly fragment:00751"
 fragment_chain:1
 15304. .17463
 /notes="assembly fragment:01245"
 fragment_chain:2
 17564. .19968
 /notes="assembly fragment:01608"
 fragment_chain:2
 20069. .24483
 /notes="assembly fragment:00176"
 fragment_chain:2
 24584. .34118
 /notes="assembly fragment:00302"
 fragment_chain:3
 34219. .43750
 /notes="assembly fragment:02455"
 fragment_chain:3
 43851. .46349
 /notes="assembly fragment:00304"
 fragment_chain:4
 46450. .49796
 /notes="assembly fragment:03676"
 fragment_chain:4
 49897. .53042
 /notes="assembly fragment:00365"
 fragment_chain:5
 53143. .56709
 /notes="assembly fragment:00090"
 fragment_chain:5
 56810. .75751
 /notes="assembly fragment:00610"
 fragment_chain:6
 75852. .98190
 /notes="assembly fragment:03362"
 fragment_chain:6
 98291. .100572
 /notes="assembly fragment:01755"
 fragment_chain:7
 100673. .109672
 /notes="assembly fragment:02371"
 fragment_chain:7
 109773. .118329

```

misc_feature /note="assembly fragment:02033
              fragment_chain:8"
              118430..123121
              /note="assembly fragment:00196
              fragment_chain:8"
              123222..136596
              /note="assembly fragment:03035
              fragment_chain:9"
              136697..139702
              /note="assembly fragment:00105
              fragment_chain:9"
              139803..148966
              /note="assembly fragment:00045"
              149067..151784
              /note="assembly fragment:00250"
              151885..154359
              /note="assembly fragment:00546"
              154460..156467
              /note="assembly fragment:00581"
              156568..160930
              misc_feature
              misc_feature
              misc_feature

Query Match
Best Local Similarity 100.0%; Pred. No. 9.9;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1395 ATTCTTTTGAAGAAATTC 1415
Db 237441 ATTCTTTTGAAGAAATTC 237461

RESULT 32
AF322045 1472 bp DNA linear INV 02-MAY-2002
LOCUS Malawimonas jakobiformis chaperonin subunit delta gene, partial
DEFINITION cds.
ACCESSION AF322045
VERSION AF322045.1 GI:18033033
KEYWORDS Malawimonas jakobiformis.
SOURCE Malawimonas jakobiformis.
ORGANISM Eukaryota; Malawimonadidae; Malawimonas.
REFERENCE 1 (bases 1 to 1472)
AUTHORS Archibald,J.M., O'Kelly,C.J. and Doollittle,W.F.
TITLE The chaperonin genes of jakobid and jakobid-like flagellates:
        implications for eukaryotic evolution
        Mol. Biol. Evol. 19 (4), 422-431 (2002)
JOURNAL 21918420
MEDLINE 11919283
PUBMED 2 (bases 1 to 1472)
REFERENCE Archibald,J.M., O'Kelly,C.J. and Doollittle,W.F.
AUTHORS Direct Submission
TITLE Submitted (17-NOV-2000) Biochemistry and Molecular Biology,
JOURNAL Dalhousie University, 5859 College St., Halifax, NS B3H 4H7, Canada
FEATURES
    source
        1..1472
        /organism="Malawimonas jakobiformis"
        /strain="ATCC 50310"
        /db_xref="taxon:50310"
        /db_xref="taxon:136089"
        join(45..134,200..343,400..519,581..715,777..902,
        965..1372,1446..1472)
        /product="chaperonin subunit delta"
        join(45..134,200..343,400..519,581..715,777..902,
        965..1372,1446..1472)
        /note="CCrDelta; Tric"
        /codon_start=1
        /product="chaperonin subunit delta"
        /protein_id="AAU56961.1"
        /db_xref="GI:18033034"
        /translation="IITPTGDIITNDGATILKQMDIQHPAGKMLVELSKSDIEAG
        GTTSVVMAGSLITVOQLKGIHPISDSVLANKIGSETLQDMAFPVTLDRRO
        LTKAVNTALSSKAHIFTSLSPAVDAVRVIDPATNVDDIKDIRVVRGLGTYDD
        TELIDGLVLPDRPNEASAKATPKLVKAKIGLIGQCISAPKTNMESNLVVDGYDM"

CDS
        mRna
        CDS

```

```

BASE COUNT 314 a 351 c 372 g 434 t 1 others
ORIGIN
Query Match
Best Local Similarity 100.0%; Pred. No. 41;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1709 GCACCACTCTGTGTGTGT 1728
Db 243 GCACCACTCTGTGTGTGT 262

RESULT 33
CAU38234 2895 bp DNA linear BCT 06-NOV-1996
LOCUS Clostridium acetobutylicum phosphotransacetylase (pta) and acetate
DEFINITION kinase (ack) genes, complete cds.
ACCESSION U38234
VERSION U38234.1 GI:1491788
KEYWORDS Clostridium acetobutylicum strain=ATCC 824.
SOURCE Clostridium acetobutylicum
ORGANISM Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
        Clostridium.
        1 (bases 1 to 2895)
        Boynton,Z.L., Bennett,G.N. and Rudolph,F.B.
        Cloning, sequencing, and expression of genes encoding
        phosphotransacetylase and acetate kinase from Clostridium
        acetobutylicum ATCC 824
        Appl. Environ. Microbiol. 62 (8), 2758-2766 (1996)
JOURNAL 96316392
MEDLINE 8702268
PUBMED 2 (bases 1 to 2895)
REFERENCE Boynton,Z.L., Bennett,G.N. and Rudolph,F.B.
AUTHORS Direct Submission
TITLE Submitted (06-OCT-1995) Biochemistry and Cell Biology, Rice
JOURNAL University, 6100 South Main, Houston, TX 77005, USA
FEATURES
    source
        1..2895
        /organism="Clostridium acetobutylicum"
        /strain="ATCC 824"
        /db_xref="taxon:1488"
        128..1129
        /gene="pta"
        128..1129
        /gene="pta"
        /note="catalyze the reaction from acetyl-coA to
        acetyl-phosphate"
        /codon_start=1
        /translation="WDLRSTWCKQKQKRIILAGEEKNLIADKIKKGLAEV
        LVGDENKIKERASELNLDISKAEIMDPETSLTEYARFYLRRHKWTIKSEKTV
        RPLPLFATNALDGVNDGAVSGAVTTGDLRPGQIIKTAPGVKIVSGFVMIIPDC
        DVGEGSLIFADCAVNPNTSDELADIAITTAETARKLCNVPKAMLSFTMGAKG
        EMDVKNAVETTKKRPRLAIDGELQDAALDSRYVALKAPSSNVAGNAVLPDDL
        QTNIGYKLVQVFAPAKAIGPICQGFAPRINDLSRSCSESDLVNVVAITVQAQKI"
        1141..2346
        /gene="ack"
        1141..2346
        /gene="ack"
        /note="catalyze the reaction from acetylphosphate to
        acetate"
        /codon_start=1
        /translation="
        /product="acetate kinase"
        /protein_id="AAI18301.1"
        /db_xref="GI:1491790"

```

```
/translation="MKVLVINGSSSIKYQFDMKDETVLAKGLVERIGKGVITHK
VNGKIVTETPMEDHKRAIKULVDALLNDEYGVIKNIIDEISAVGHRIVHGGEKYANSV
LIDBVKVSEDCVSLAPLHNPPIIGINACKELMPNVPVAVFDTAFHOTIPDYAYM
YAIPEYDYDKTKRYGFHGTSHRYKVSRTAAEFKGVKVEDLKVVCHMGNGASITAVE
NGKSVDSMTGTPPGGLAMGARGSDMDPAVTFILMDKLININASVNNLNKSGIEGL
SGISSMEDIKKGNVVDKPKAMLAYSVFTVKIKQFSGSVAVNGLDCIYFTGGIGE
NSPFERREICKMMDYLGIKIDDKQNDGTMGIPMDISAGSKRVVLVPTNEELMIARD
TKDVGKDK"
2422..->2895
/note="ORF X"
/codon_start=1
/translate=1
/product="unknown"
/protein_id="AAB18302.1"
/db_xref="GI:1491791"
/translation="MLDVSDLITKKVKDKQVDIQLGMYTFYDSEYKILEPLTKGK
LLMDSITLVAEYKIVELVETCSLKQKFPYDIDLEIKFTDPEKNKDDVEFINNY
EVDINEIVENNIILSLPIKKLCRDCKGLCPMCGTNLISNCNCHDNIDPRFAKL"
BASE COUNT 1110 a 354 c 576 g 855 t
ORIGIN
Query Match 1.1%; Score 20; DB 1; Length 2895;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 480 ATGCAATGTTCTTGATTC 499
|||||
Db 927 ATGCAATGTTCTTGATTC 946
|||||
RESULT 34
IGU48714/c
LOCUS
DEFINITION Idiomylia grimshawi alcohol dehydrogenase (Adh) gene, complete cds.
ACCESSION U48714
VERSION U48714.1 GI:1256247
KEYWORDS
SOURCE Drosophila grimshawi.
ORGANISM Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 3644)
Brennan,M.D., Thorpe,P.A., Hu,J. and Dickinson,W.J.
Molecular organization of the alcohol dehydrogenase loci of
Drosophila grimshawi and Drosophila hawaiiensis
Gene 181 (1-2), 51-55 (1996)
97128767
8973307
2 (bases 1 to 3644)
Brennan,M.D.
Direct Submission
Submitted (06-FEB-1996) Mark D. Brennan, Biochemistry, University
of Louisville, 319 Abraham Flexner Way, Louisville, KY 40292, USA
Location/Qualifiers
1. 3644
/organism="Drosophila grimshawi"
/strain="G1"
/db_xref="taxon:7222"
2116..3047
/gene="Adh"
join(2116..2208,2303..2707,2781..3047)
/codon_start=1
/product="alcohol dehydrogenase"
/protein_id="AAB40357.1"
/db_xref="GI:1256248"
/translation="MVJANGNIIFVAGLGIGLTSREIVKSGPKNLVLDVDPNPA
IAELKLNPKVTITFYFDVTVPLAETKKLLKTIKFKLTVLDLLINGAGILLDDNQIER
TIAVNTGTVNTTITAILDFWDKRGKGGVAVNICSVTGENSIYQVPVYSKAAALS
FTTSIAKLHITGTAVSINGPITKTVLHKNSWLGVPRVAELLLHEHTQTTLQCA
QNFYKAEANGQAIKWLGLRGLDAIEWTKHWDGSI"
BASE COUNT 1230 a 682 c 645 g 1087 t
ORIGIN
Query Match 1.1%; Score 20; DB 5; Length 7370;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 891 TTTTGGGAATCGATTTTAAA 910
|||||
Db 3103 TTTTGGGAATCGATTTTAAA 3084
|||||
RESULT 35
AF043383
LOCUS
DEFINITION Pseudopleuronectes americanus amino-peptidase N (ampN) gene, partial cds.
ACCESSION AF043383
VERSION AF043383.1 GI:3452274
KEYWORDS
SOURCE Pseudopleuronectes americanus.
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Pleuronectiformes;
Pleuronectoidei; Pleuronectidae; Pseudopleuronectes.
1 (bases 1 to 7370)
Douglas,S.E., Gallant,J.W. and Bullerwell,C.E.
Molecular Investigation of Amino-peptidase N Expression in the
Winter Flounder, Pleuronectes americanus
J. Appl. Ichthyol. (1998) In press
2 (bases 1 to 7370)
Douglas,S.E., Gallant,J.W. and Bullerwell,C.E.
Direct Submission
Submitted (15-JAN-1998) Marine Biology, Institute for Marine
Biosciences, 1411 Oxford Street, Halifax, NS B0J 1G0, Canada
Location/Qualifiers
1. 7370
/organism="Pseudopleuronectes americanus"
/db_xref="taxon:8265"
<1658..>7370
/gene="ampN"
join(<1658..1988,2897..3203,3500..3642,3736..3872,
4142..4268,4483..4637,5129..5242,5364..5507,5744..5809,
6088..6153,6231..6403,6566..6642,6873..7006,7098..7153,
7361..>7370)
/gene="ampN"
/product="amino-peptidase N"
join(1658..1988,2897..3203,3500..3642,3736..3872,
4142..4268,4483..4637,5129..5242,5364..5507,5744..5809,
6088..6153,6231..6403,6566..6642,6873..7006,7098..7153,
7361..>7370)
/gene="ampN"
/EC number="3.4.11.2"
/codon_start=1
/product="amino-peptidase N"
/protein_id="AAC32807.1"
/db_xref="GI:3452275"
/translation="MGKGFYISKAVGVGVILGAGALATIIALSVVYSQEKAKNNNE
VSTVGTPTGPTVPTTSPAPVTPAPSNAPKPYLPKSLVPHVNVNKLWPLRTADP
TTGLYFTGESTVEFKCTEDTLILHSKNLTNLTDLNGQWALSAVNSGVKAPAIT
SRLOPVTYLVLDQGLKLLQWYHLFTDFTGLADDDLGFGFYRSVYVNEGORKVAVT
QMQPTDARKAPFCDFEPAKMATFTNITLIHDPETVALSNGAQRSPKVTIDGNLQKTD
FEQTERKSTYLLAFIVSEFTSINNTVDNVLIRIPARKPAIDAGQAGAYLSKTPILKF
IEHELAMFGLNVLTRWMDLNLNEGFSASYVYLGADVAEPDWNIKDLIVLNDHVRV
FAYDALASSHPLSAKEDIQRPQISELFPDAISYSGASVIRMLSDFLTEDIFVNGUR
TYLKFAPFAGNAVYTDLWNHQLQMAVNTGTLPGSVQDMMTVLWLMQMGFPVPTVITSG
EVSKQFLDLPDSEVPNTSKWIKVIPTKTATQAQPPYVLEQSKATNKKTTGV
DWTLANLVVGYRVNYDSSNWDKLLKLVSTNHLQIVINRAQLVDDAFNLARAKI"
BASE COUNT 2197 a 1546 c 1483 g 2144 t
ORIGIN
Query Match 1.1%; Score 20; DB 5; Length 7370;
Best Local Similarity 100.0%; Pred. No. 40;
```


TOYSMIAIEKESBEQVONPLMIVRNHALEGASLKNTRPEFAVGEVGVNFGVXDKATF
SDRETPVQGFHWKSPKCGEVSCTKNGLPFGSGRWRMAVLDGVVISPVLNVLK
DHASVGNFYREAVHTAKDLKSGAMSFPEILSEEVVPELSSORVOGLISVLGL
NVLIVMSVYVYRCGVVTAIVLNLILLIWSMOYLDAPLTLSCLAGIILAMGAVDA
NVLIVFERIREYLLTSLSSVBAGYKAFSAIPDSNLTTILASALLMLDGTPIKGF
ALTLLIGIFSSMTALPMTKFFVFWQKTRQTLHMMKFIIGIKHNFLECKRKLWV
SGVVLGCVGLGAGWDSVFGWDFKGGYALTLDSDVCYNPEQMCVLRKFOQJGL
SSRDYRVRADSEKVIYLSQNALDRVEQIEGAGSEQKSDHLARVLQVLSDSGS
TTSWVPDASRGWFKVSGQSNKRWRTQAVIALFGALGIIILFYVSLRFEWRYAFAICA
LMHDLATCAVLVAFHFQRIQIDLOALGALMTVLGYSLNLTLLIPDRIRERREKL
FTWPIPLINDALOKTGRVTWTTATTLVSLVILLFVGGSGIFNFAFIMTVMGILLGLTS
SLVIAPLLLMVFKERQNSLR"

BASE COUNT 3118 a 2399 c 1722 g 2750 t
ORIGIN

Query Match 1.1%; Score 20; DB 1; Length 9989;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 761 TTGCATATGATGAATAAGTT 780
|||||
Db 6559 TTGCATATGATGAATAAGTT 6540
|||||

RESULT 37
AE007683
LOCUS
DEFINITION Clostridium acetobutylicum ATCC824 section 171 of 356 of the complete genome.
ACCESSION AE007683
VERSION AE007683.1 GI:15024703
KEYWORDS
SOURCE Clostridium acetobutylicum
ORGANISM Clostridium acetobutylicum
Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae; Clostridium;
1 (bases 1 to 10742)
Nolling,J., Breton,G., Omelchenko,M.V., Markarova,K.S., Zeng,Q., Gibson,R., Lee,H.M., Dubois,J., Qiu,D., Hitti,J., Wolf,Y.I., Tatusov,R.L., Sabathe,P., Doucette-Stamm,L., Soucaille,P., Daly,M.J., Bennett,G.N., Koonin,E.V. and Smith,D.R.
Genome sequence and comparative analysis of the solvent-producing bacterium Clostridium acetobutylicum
J. Bacteriol. 183 (16), 4823-4838 (2001)
21359325
11466286
2 (bases 1 to 10742)
Childress,D., Zeng,Q. and Smith,D.R.
Direct Submission
Submitted (24-JUL-2001) GTC Sequencing Center Production, Finishing, and Bioinformatics teams, Genome Therapeutics Corp., 100 Beaver Street, Waltham, MA 02453-8443, USA
Location/Qualifiers
1..10742
/organism="Clostridium acetobutylicum"
/strain="ATCC 824"
/db_xref="ATCC:824"
/db_xref="taxon:1488"
111..2147
/gene="CAC1736"
/note="recG"
111..2147
/gene="CAC1736"
/codon_start=1
/transl_table=11
/product="RecG helicase"
/protein_id="AAK79702.1"
/db_xref="GI:15024704"
translation="MNINDDIATLKGVPKTLKXKGIPTILDLLLYPRDYENLS
LMSNILEVDQKIVKCTPLRLEKFRSKRTVTKIIFSDGKTTFGIWFNOPYIK
NKFYNTYITMGIKETKNEITMNPPIPVENRSSDEKIVPKVLSGTLKNTFFIRL
IFELLQNIKIDNPEWIKKNAPHSDESUREIHSFPMLELRESTRRLKPKQLFYI
SLKVLMLKYEIKKNKGFAPKFIAPELIKLKESLPFKUTEAQNULVIREILKDEKRPQAM

NRLVOGDSGKTIIVALIAIPNVKNGFOAVLMAPTEILAKOHFESANTLKFKEFNIRI
ELLTSTVDKNKRIIKEKALDEIDLIICHTHALIEDNVSFSLNIGIIVVDELHFRGVMQ
RRLFNKGNIIIDVLWTATPIRTLLALYLDGDDLIIDTLPGRKEVTKLCTCKNKR
QAYNPALEIKNGRIQYVVCVEENKELETSVEKLYEELKSYFKGIKLEMLHKG
MAPKENAMERFNKNTETVLSTVIEGVVNVNASVMIIEDAQRFGLAQHQLRGR
VGRGQYOSVCLVAEMKNKVAERMEIMCSDSGDFYISBEDFKLRGSGEVFGIRQSGE
DGLILSDIIDIISILREANSAAKKLIISTEKSDIVIKKSVKSLKNTSKYICPN"
2225..2791
/gene="CAC1737"
2225..2791
/gene="CAC1737"
/codon_start=1
/transl_table=11
/product="N6-adenine-specific methylase"
/protein_id="AAK79703.1"
/db_xref="GI:15024705"
translation="MRIISGKAKRKILPPKGMETTRPTLDRVKEAMFNIONDVPEA
VLDMPFGTSGLSGLEAASGAKVCYLIDKSPITYPILKENVENLRDECKTNMDSY
EAVYILASKGKEFTLIFIDPPYAKEMIIPPAIDLICKEKULTKSGLIVTKIDSDEKIYE
GNNEIIVDHRKYKTIIVCFYRYKEDK"
2788..3273
/gene="CAC1738"
/note="kdtB"
2788..3273
/gene="CAC1738"
/codon_start=1
/transl_table=11
/product="Phosphopantetheine adenylyltransferase"
/protein_id="AAK79704.1"
/db_xref="GI:15024706"
translation="MKVAVYPGSFDPTNGHLDLIISRAKVFQKVIUGVLINPEKKGM
FSVDVLDIKRVIKPNNVVSQSGLLVNFNENNSNMIRGLRASVGDFEVLQTS
LNKKLNPNVETVFMNTSLSEFSLSTAIKQAVFGCTIKELVPDEIIDDLVLRKAEF
N"
3345..3869
/gene="CAC1739"
3345..3869
/gene="CAC1739"
/codon_start=1
/transl_table=11
/product="Uncharacterized conserved protein (coiled-coil)"
/protein_id="AAK79705.1"
/db_xref="GI:15024707"
translation="MNVVKLLEYLGDIVDTSSKVPLSKNVKNVKNKEILDVITQMSEM
PEELKQWILSEKDLINDAVERNEIKKQETETIKMDKHSITLEAEKRAKRLLE
DAENAKQIKLSALHYVDELLCHLDSQLDROQFNEKQVREASQLSINVQDNFSIR
EIKNNISREYVK"
complement(3862..5007)
/gene="CAC1740"
complement(3862..5007)
/gene="CAC1740"
/codon_start=1
/transl_table=11
/product="Uncharacterized membrane protein, YLBj
B.subtilis homolog"
/protein_id="AAK79706.1"
/db_xref="GI:15024708"
translation="MIIFLLIIIFLITALLFLLIKELHLSNASNYIMTIFLSLIILF
IILKPEVCISASISGAKLPFPVPVPEPLITNLICVDGIEIVSKLLGILCLPO
RLPKVASIVLWVSLGYPGLGAKYAADLINDIISFTFELVNIASNPGLFIIGAV
GTSMLKNNKYLIGYLLISCYLSCIAMGIIIPNKKDDFTKTKWNSKHSIGTAKTS
TENAVKILQWMSFIIIFSVIIGILKQSVIFKNSSEAFIKTFILGIVEMTNGSSMLLS
SSPFIELKVTITISFLSSFGGLCVIAQIYSFIKYSILKFIKLVIGVQGVSSICFL
LPKPLVYVNTISTFNANNFSLSPSHFTVAIVLAVLIPILYIKKTKTI"
complement(5072..6280)
/gene="CAC1741"
complement(5072..6280)
/gene="CAC1741"
/codon_start=1
/transl_table=11
/product="Uncharacterized conserved protein, YLBM
B.subtilis ortholog"
/protein_id="AAK79707.1"

```
/db_xref="GI:15024709"
/translacion="KNITGIIAEYFNPANHGIHHEKTEICKSDVILVWAGDPVOR
GEPALIDKMSRAVALSGVDVITELPCVYSSSAEFAYGAVSLNLSIGSVNICFG
SERGEHMD.YLISKILVDEPYEYSILKQYLNQGFSPRMEALMEYNNLSIKVDH
GNDHLSISSNIGVEYCKSLIKLSKNIPYTIKREGNMYNDLEIKTISASAIANA
LKSKEKISLRYOVPKKTFDILSDNINSICYKOYIPDYIKYALTSKEINLKPVE
GIDNKTYNLLKCSNMDSLMTITKMYRYYSRSRIILTOYFGDCVYVESIRNSPC
VARIIGFNKKGKIALKEFKTSSIPILITVNNYNPDLSDINATAYSLINKSVNPL
DDYRKIIIV"
/gene="7546"
/gene="CAC1742"
/notes="pica"
/gene="7546"
/gene="CAC1742"
/codon_start=1
/transl_table=1
/product="phosphotransacetylase"
/protein_id="AAK79708.1"
/db_xref="GI:15024710"
/translacion="MDLISIMECAKODKRRIILAEGERKNIADKTIKESGLAVI
LVGDENKIKERKSELNLDISKAIMDPEISLKTETARDFELRKHKMTIEKSEKXV
RDLVYFATMALKQYVDGWSGAVHTGDLRLGLOIITAPGVKIVSGFPVMIIDC
DYGEGLL.PADCAVBNPTSDLAITTAETARLGNVEPKVAMLSFTSGSAGK
EMWDRYKNAVEITKRRPDLAIDGLOLAIDAISEVALKAPSNVAGNAVLPEDL
QTNIGIKLVQRPAAKAKAIGPICQFAPKPLINDSRGSSSDIVNVAILTVQAORCI"
/gene="8763"
/gene="CAC1743"
/notes="aska"
/gene="8763"
/gene="CAC1743"
/codon_start=1
/transl_table=1
/product="Acetate kinase"
/protein_id="AAK79709.1"
/db_xref="GI:15024711"
/translacion="MKNIYVNGSSSIKKYOFIDMKDETVALKGLVERIGKSVITKH
VNGEKVYITPMEDHKKATKLVLDALINDRYGVIKINIDEISAVGHAIYAGGEYKASV
LIDEDVKSIEDCVSLAPLNPPIHIGINACEKLMNVEPVAVFDIAFHOTIPIAVIM
YAIPEYEDYFKYIKRYGFGHTSHRYVSRTAEBIIGKVEDLKXVAVCHMGASITVAE
NGKSVDTSMGFTPLGLAMGTRSGLMDPAVITPLMKLNNINAEVNNHNGASIGIEL
SGISDMRJCICKKNYVDKPKAMLAIVSEVNYKIKOFISYTAVMGDLCLVFGGIGE
NSFENREICCKNMDYIGIKIDDKKDNBTMGIPMDIASBSKXAVLVIPINHELMIARD
TKDIVGKX"
/gene="9387"
/gene="CAC1744"
/gene="9387"
/gene="CAC1744"
/codon_start=1
/transl_table=1
/product="Predicted Zn-finger-like protein, possible
nucleic acid binding"
/protein_id="AAK79710.1"
/db_xref="GI:15024712"
/translacion="MIMTLKNRFLRPLTKGLMDNSIIPYAEVKGIVELTSGRC
LQFPYDILDEKFTDNPENKODEVITFINNYEVINDINEVENNIIISLPKLCED
CKGICPMOCJTNLISITCNCHEDNIDPRFALKDLFSNH"
/gene="9384"
/gene="CAC1745"
/notes="trmf"
/gene="9384"
/gene="CAC1745"
/codon_start=1
/transl_table=1
/product="Ribosomal protein L32"
/protein_id="AAK79711.1"
/db_xref="GI:15024713"
/translacion="WGNPARKFSKARRDSRAQTFKSLDPGMTECPNCHMKLAHYC
KKGYYGKEIVAVEK"
/gene="10659"
/gene="CAC1746"
/notes="plex"
/gene="10659"
/gene="CAC1746"
CDS
```

```
/codon_start=1
/transl_table=1
/product="fatty acid/phospholipid biosynthesis enzyme"
/protein_id="AAK79712.1"
/db_xref="GI:15024714"
/translacion="MIIAIDMGGNDAEAVKGCVEAIKGNSSIHIIITPSDKISA
ELKKTFFBESIEIYDAKDVITNNEHPMAVARRKSESLYKALNIVKGRADAVISAG
Query Match 1.1%; Score 20; DB 1; Length 10742;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 480 ATGCAGATGCTCTGTATTC 499
Db 7344 ATGCAGATGCTCTGTATTC 7363
RESULT 38
AE014001
LOCUS AE014001 10836 bp DNA linear BCT 26-JUL-2002
DEFINITION Yersinia pestis KIM section 401 of 415 of the complete genome.
ACCESSION AE014001 AE009552
VERSION AE014001.1 GI:21960971
KEYWORDS Yersinia pestis KIM.
SOURCE Yersinia pestis KIM.
ORGANISM Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Yersinia.
REFERENCE 1 (bases 1 to 10836)
AUTHORS Deng,W., Burland,V., Plunkett,G. III, Boutin,A., Mayhew,G.F.,
Liss,P., Perna,N.T., Rose,D.J., Mau,B., Zhou,S., Schwartz,D.C.,
Fetherston,J.D., Lindler,L.E., Brubaker,R.R., Plana,G.V.,
Straley,S.C., McDonough,K.A., Nilles,M.L., Matson,J.S.,
Blattner,F.R. and Perry,R.D.
GENBANK J. Bacteriol. 184 (16), 4601-4611 (2002)
12142430
TITLE 2 (bases 1 to 10836)
JOURNAL Deng,W., Burland,V., Plunkett,G. III, Boutin,A., Mayhew,G.F.,
PUMED Liss,P., Perna,N.T., Rose,D.J., Mau,B., Zhou,S., Schwartz,D.C.,
Fetherston,J.D., Lindler,L.E., Brubaker,R.R., Plana,G.V.,
Straley,S.C., McDonough,K.A., Nilles,M.L., Matson,J.S.,
Blattner,F.R. and Perry,R.D.
DIRECT SUBMISSION Direct Submission
JOURNAL Submitted (21-FEB-2002) Genetics, University of Wisconsin, 445
Henry Mall, Madison, WI 53706, USA
FEATURES
source
1. 10836
/organism="Yersinia pestis KIM"
/strain="KIM"
/db_xref="taxon:187410"
complement(77..1066)
/gene="y3968"
complement(77..1066)
/gene="y3968"
/function="unknown"
/notes="residues 6 to 246 of 329 are 33.74 pct identical to
residues 7 to 248 of 329 from Genpept :
>gb|AAD37770.1|AF146532.10 (AF146532) glycosyltransferase
[Klebsiella pneumoniae]"
/codon_start=1
/transl_table=1
/product="hypothetical"
/protein_id="AAK7512.1"
/db_xref="GI:21960972"
/translacion="MNSVDISYIVVNCANYLTDLDSLRSGTSGVSEIIVANDGST
DHSILAEQVAHDSRLVYINQNGLSRARVGIARAKGMVVAADGDMDLAPHALS
TWLOVKQKQKLDVILGNGYQFSNPERKVPVPLCQVPGATILNQOMITISYKQKEW
PHYAWLQILRVDVISTHQLQFTIRMLHEDITLTLQALVAQRIGCATPFGYRTNPD
SITNSPMQALIRANSYVDIINALVKIAASTKPPVRDALSHAOQESGHLGLRKK
LPASPORTALAIRFTILGLRPALFRGVNTHLMRAIRCSFLMGSAVQKRAE"
complement(1201..2586)
/gene="y3969"
gene
```


gene	complement(1201. .2586) /gene="y3969" /function="putative transport" /note="residues 1 to 455 of 461 are 79.34 pct identical to residues 180 to 633 of 637 from E. coli K12 : B3352; residues 1 to 450 of 461 are 80.66 pct identical to residues 180 to 629 of 635 from GenPept : >emb CAD08154.1 (AL627281) probable ABC transporter ATP-binding protein [Salmonella enterica subsp. enterica serovar Typhi]" /codon_start=1 /transl_table=11 /product="putative ATP-binding component of a transport system" /protein_id="AA087513.1" /db_xref="GI:21960973"
CDS	/translation="MDLDAVILWLEKWLKSYTGLVLSHSDRDLFDPIIDKILHIEQQT LNEYTGYSSFERQATKLSQQSQMYHQQKVAHLQSYIDRFRAQATKAAQASRIK MLERMLIAPAHVDNPHFSPFTESLPDPLLRMDKVSAGYGVNTVLSQIKLNVPGS RIGLLGRNGACKSTLILKLAGTLPQSCEIALSKIGLGYAQHQLERLADSFLOQH MSRLAPKSEQLRDYLGGYFGQDQVDTPTARFSGGEKARLVLAIIWQRPNLLLD EPTNHLIDMDKQALTEALIDFEGALVVVSHDRHLRSTDDLYLVHDKGVQFPDGLD DYQWLVDQSQQQDNQPAKELSGNSAQQRDKQRDAEFRTQTQPLRKQIMTLEKQ MDKLSLELAIAEQLADSALYDIARKADLTQCLLQQTQVKSLEBETEMQWLDVQEQLE ILTKFESDADI" complement(2547. .3119) /gene="y3970" complement(2547. .3119) /gene="y3970" /function="transport" /note="N-terminal fragment of ATP-binding protein of ABC transporter; naturally broken Ofc; residues 1 to 187 of 190 are 80.74 pct identical to residues 1 to 187 of 637 from GenPept : >gb AAG58460.1 AE005558.13 (AE005558) putative ATP-binding component of a transport system [Escherichia coli O157:H7 EDL933]" /codon_start=1 /transl_table=11 /product="hypothetical" /protein_id="AA087514.1" /db_xref="GI:21960974"
gene	/translation="MIVFSSLIQIRRGTRVLDDNATINPGQKVLGKNGCGKSTLL ALLKGLSADGNTVFPNNMALWNOETPALDIPATIEYVIDGDEYRQLEAELOAAN EKNGHATATVHGKLDIAHANTIQSRAASLLHGLFGSQEKLQPPVRSFGSGWRRLNL AQALVCRSDLLLDLEPTTWIMWPSGKNG" 3367. .3918 /gene="y3971" 3367. .3918 /gene="y3971" /function="putative" /note="residues 1 to 182 of 183 are 75.27 pct identical to residues 1 to 182 of 184 from E. coli K12 : B3351" /codon_start=1 /transl_table=11 /product="putative NAD(P)H oxidoreductase" /protein_id="AA087515.1" /db_xref="GI:21960975"
CDS	/translation="MMLQPPKVLVLYAHPEQSQSVANKRVLLQPVQQLHVTVDLYAH YPDFIDIHFEQQLIRHQVIVFQHPLYTSCPALLKEWLDRVLARGFANGVGGHALT GKHWSRVTTGQECTYRIGGNYRYPMEDILRPELTAAMCHWINPMIYIYWARQK PETLASHAAQVQLQSLTRGL" 3922. .5730 /gene="kefB" /note="y3972" 3922. .5730 /gene="kefB" /function="transport; transport of small molecules; catons" /note="residues 1 to 602 of 602 are 77.90 pct identical to residues 1 to 601 of 601 from E. coli K12 : B3350" /codon_start=1 /transl_table=11 /product="K+ efflux; NEM-activable K+/H+ antiporter" /protein_id="AA087516.1"
gene	/db_xref="GI:21960976" /translation="MEGTGLLTAVLVFELFAVAVPIAQRIGIGAVGLIAGIAGIP WGLFIRDVDEILHFSVFLFIIGLELPAKLQWLRRSIFGVGAGQVLLTAAVL GALLYFTQFAWAQAVIGGVGLAMSSTAMALQMKRKNRNEGQLGFSVLLFODMAV IPALALPILAGNEGANDWKIGLKIATAAPAGMLIGGRYLLRLPLFRIVASGVREVFT AALLVVLGSALEMDALGSLMAGTFTIAGILLAESEFQHELEIAIEPFGILLGLFFI SVGMALDGLVFTLLDVLGLVLAIVFIKSAIILGLARVGLRRSVLRQFAGVUSQGG EFAPLPSAAPSQRVLNAEQALLLVVTLSMMTTPLLMQVIDRILLRVYNAQESDE KPFVEDNDPQVIVGFGFQVGRLLMANKRITVLERDVSAYSMRKYGYKYVYGD ATELELLRAAEAKAIVITCNEPEDTMALVHLCCQHPFNHLILARARGVEAHELL QNGVKDFTRETFSALBGRKTLLELGMHPHOAVRAQCHPRLDMRLRELMPQHGD VAQISRIKEARRELEDFQREMLHESQLDGDWEYE" 5746. .5967 /gene="y3973" 5746. .5967 /gene="y3973" /function="unknown" /note="residues 11 to 73 of 73 are 65.07 pct identical to residues 4 to 66 of 66 from GenPept : >gb AAL22319.1 typhimurium LT2]" /codon_start=1 /transl_table=11 /product="hypothetical" /protein_id="AA087517.1" /db_xref="GI:21960977"
CDS	/translation="MTITLNTKTVTRKFIAGAVCPQCNALDTLALWREDQVVEVCVK CGHROQTDEQVKNKVRPHEQVIGIFPPE" complement(5836. .6063) /gene="y3974" complement(5836. .6063) /gene="y3974" /function="unknown" /note="residues 15 to 70 of 75 are 26.78 pct identical to residues 112 to 167 of 1024 from GenPept : >emb CAA53232.1 (X75544) beta subunit RNA polymerase [Plasmodium falciparum]" /codon_start=1 /transl_table=11 /product="hypothetical" /protein_id="AA087518.1" /db_xref="GI:21960978"
gene	/translation="MISPTLGTIGTDCGNLPRCTHPHKKSROIITLLWRKNTDHLML WTNMFIDFLRLALMTTLNTHYFNLIPTP" 6062. .6649 /gene="slyD" 6062. .6649 /note="y3975" /gene="slyD" /function="enzyme; proteins - translation and modification" /note="residues 1 to 195 of 195 are 83.41 pct identical to residues 1 to 196 of 196 from E. coli K12 : B3349; residues 1 to 195 are 82.91 pct identical to residues 1 to 196 of 196 from GenPept : >gb AAL22318.1 (AE008858) FKBP-type peptidyl prolyl cis-trans isomerase (rotamase) [Salmonella typhimurium LT2]" /codon_start=1 /transl_table=11 /product="FKBP-type peptidyl-prolyl cis-trans isomerase (rotamase)" /protein_id="AA087519.1" /db_xref="GI:21960979"
CDS	/translation="MKVTKDLVLSLAYQVTEGVLVDSPVSAPLDYLHGHSIAG LENALEHAGDSFDVRYNADEGYSYDENLVQRVQKDVPMGVDELVGRHFLADTDQ GPVPEITAVEDESHVVGDVGNHMLAGQDLNFHVVAVREATEBELQHGHHGEHDHHH EHGDCGCGHGHDDHEHEHGGKCGCGCGCGCH" complement(6870. .7088) /gene="slyX" /note="y3976" complement(6870. .7088) /gene="slyX" /note="residues 1 to 72 of 72 are 69.44 pct identical to

Query Match 1.1%; Score 20; DB 1; Length 10836;
Best Local Similarity 100.0%; Pred.No.39;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 304 TGGCTTGGCAATGCTTATG 323
|||||
Db 8567 TGGCTTGGCAATGCTTATG 8566

RESULT 39
AE002341/c 12442 bp DNA linear BCF 26-MAY-2000

LOCUS Chlamydia muridarum, section 69 of 85 of the complete genome.
DEFINITION AE002341 AE002160
VERSION AE002341.2 GI:8163305
KEYWORDS
SOURCE Chlamydia muridarum.
ORGANISM Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.

REFERENCE 1 (bases 1 to 12442)
Read,T.D., Brunham,R., Shen,C., Gill,S.R., Heidelberg,J.F.,
White,O., Hickey,H.K., Peterson,J., Umayam,L.A., Uetreback,T.,
Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B.,
Bowman,C., Dodson,R., Gwin,M., Nelson,W., DeBoy,R., Kolonay,J.,
McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.
Genome sequences of Chlamydia trachomatis Mohn and Chlamydia
pneumoniae AR39

TITLE Nucleic Acids Res. 28 (6), 1397-1406 (2000)

JOURNAL MEDLINE
20150255
PUBMED 10684935

REFERENCE 2 (bases 1 to 12442)
Read,T.D., Brunham,R., Shen,C., Gill,S.R., Heidelberg,J.F.,
White,O., Hickey,H.K., Peterson,J., Umayam,L.A., Uetreback,T.,
Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B.,
Bowman,C., Dodson,R., Gwin,M., Nelson,W., DeBoy,R., Kolonay,J.,
McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.
Direct Submission
Submitted (01-MAR-2000) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
On Jun 1, 2000 this sequence version replaced gi:7190754.

TITLE Location/Qualifiers

COMMENT
FEATURES
source
1..12442
/organism="Chlamydia muridarum"
/db_xref="taxon:83560"
/note="s/nonym: Chlamydia trachomatis Mohn"
complement (86..544)
/gene="TC0726"
complement (86..544)
/gene="TC0726"
/note="similar to SP:P18587; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="15 kDa cysteine-rich outer membrane protein"
/protein_id="AA39536.1"
/db_xref="GI:7190755"
/transl_table="MSTTPVSGVTSQNNSSENVNNAASLTLLKERASKILSTARKV
GLAVGIFAVILSIVFLIPATPAANPYLAIPALGCVNICIGILSNMKGSCSEAK
WKLCKNVLTSESDILDDGELNNSNKIFTDNLRSRVEDIYITLSSRNSVA"
complement (727..2391)
/gene="TC0727"
complement (727..2391)
/note="similar to ; identified by sequence similarity;
putative"
/codon_start=1
/transl_table=11
/product="60 kDa outer membrane protein"
/protein_id="AA39537.1"
/db_xref="GI:7190756"
/transl_table="MRIGDPMNKLIRAVTIFAVTSVASLGVLETSMAESLSTNV
ISLADTKAKETTSQKDKARKKHQKRTSVVRKEVTAADVTKAVRPRDSCKQKMTYV

gene
CDS
complement (2539..2805)
/gene="TC0728"
complement (2539..2805)
/note="similar to ; identified by sequence similarity;
putative"
/codon_start=1
/transl_table=11
/product="cysteine rich outer membrane protein 3"
/protein_id="AA39538.1"
/db_xref="GI:7190757"
/transl_table="MKRTALLAALCSVSLSSCCRIYDCCEDPCAPIKSPCKSKR
EVNGGNSGNGVPSKCPGGBLDEHKQGFQARCIQADGRCRO"
3133..3357
/gene="TC0729"
3133..3357
/gene="TC0729"
/note="conserved hypothetical protein; identified by
glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AA39539.1"
/db_xref="GI:7190758"
/transl_table="MKRPFLLMATAALLTGCCSYEPKREYIILAQRCSPLFKIAQAI
NFSLQAQAEKTEFTVTVPSSLNKLSELSQN"
complement (3354..4874)
/gene="TC0730"
complement (3354..4874)
/gene="TC0730"
/note="similar to PID:1783380; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="glutamyI-tRNA synthetase"
/protein_id="AA39540.1"
/db_xref="GI:7190759"
/transl_table="WTFQNVVRAVAPSPGDPHVGTAYMALENEVFAKFNKGMLIRI
EDTDGTRSRDYEKNI FSAIKMVCGRMDEGPDIGPYGYPYOSERTETIYKXAEILQ
TDYAKCFATPOELOEMRAVASTLGVRGYRRXYLSAEVRORBOGOPYTIRLKV
PLTGESVPEDDCKGRVVPMAVDVDDVYKSDGPTVHFRAVVDHMLGITHVLRGEE
MLSTPRKULIYEARGWEPPOFFHNPILLNDGSLSKRNKPTSTFYRDAQYKKEAF
NMFLLTGWISMEGDEILYSQRLIEAFPKIIGRGAVFDIRKLDMMKXHYNHGSP
ESLKEILDMWLNDEFILKILPLCQTRITTLADFLRLTSFFMALPEYSKDDLPSLSL
GHQQAAMVLYSLVYKLEKDDMEKDPFYQSGKMLAEAVQVHKKAVIPLLVYITIGEK
GGLPLFDSMELLKARTRARLTHAQNLLGVPKRLQOQIDVALQOQPLEIRFLDF"
complement (5124..5681)
/gene="TC0731"
complement (5124..5681)
/gene="TC0731"
/note="similar to GB:113598 SP:006566 PID:289835;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="eno protein"
/protein_id="AA39541.1"
/db_xref="GI:7190760"
/transl_table="MECQODTMAVEVQEOOEENAVANARVITOAKLHNV
TRQATVYAIKOKKLKASKSTRMEIDLOLEBYRRRYSRAKSTGGLLPNEKGFYS
VQVAVSMANVEQKIYATRIKAMKGERGSAWIHSEVDRYRDYIKKAERKGN
LSAMKEGFTIGASELLADSEDFVS"
complement (6143..7891)
/gene="TC0732"
complement (6143..7891)
/gene="TC0732"

```
/note="similar to GB:M54884 SP:P21893 PID:147550
PID:897842 GB:U00096; identified by sequence similarity;
putative"
/codon_start=1
/transl_table=1
/product="single-stranded-DNA-specific exonuclease RecJ"
/protein_id="AAF39542.1"
/db_xref="GI:7190761"
/translation="MEKDKLSALQKQWYPKHNEEFLLKKIKELYLHPVIAQVLVSRG
FQSVQEDRLYQFPLSHPASSFLDMKAVARLLKSKENNEHVMYIGDGDVDGITGV
TLVVEFGQVTKVSCSGTLPKQHEITSLISOMLQDQVSLITVDCGTTAGREVR
SINKQGDVIVTDHTTGTGKLPKHCATLNPCLDKNSYPNKELTGVGAVKLVCAVTE
LIXQDASKDKIDILHLDLVLSTIADVGRSGENRILVSGIKETKGRGLKGLK
CSLSQKSDISSNLTGIRITPKNSLGRADSSQGVKLLSQDPRVTGIVTSLAIV
NQERQAEVLRVDERILANPKLTAQSAIVLASPYWHSRVPIIISARLARTKPV
AIALQDGKSGDEATLPLDVCMDQFQISODLISSMELLEPPFGKGNPSPVFTYRQ
FIHLVNSLGRKDEATLPLDVCMDQFQISODLISSMELLEPPFGKGNPSPVFTYRQ
HYVPLKSGNHVLYLNSGERNLGTAGFGQDKISLLKANNMELDIATYLRVMSRV
HGAVRLLIQDPRIQAP"
complement(8009..12211)
/gene="TC0733"
complement(8009..12211)
/gene="TC0733"
/note="similar to GB:AL009126; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=1
/product="secDF protein, putative"
/protein_id="AAF39543.1"
/db_xref="GI:7190762"
/translation="MKRNFKNLSVIFVAFALYHVLPTCLYYSRPLNKKIEGREAH
KILRLTQQADARNLUTVRKRLSTKLKRGSVKPHLSIPGVNVNHFPLAADAARFI
DNVHGSHVPTKSAIRLYLGSSELTGETIVQVSGSLVTLTEEDFSVYEDKLVV
SOELNATLFLNVNVECSCHGYSSLDMPKIEKVAHARGICSLGLLTLKTEQAFL
SRFLGTQDFVAFRLVRENALHEDVKEYEELNEAAVLLHSLRNWNEMS PRVGVN
VDCSELSPCSSSLTGDGKLLHFDADVLAKQQLGDERLRLFAVEKQHIATR
LNRSVETESGPAFLRDEAASGKIVLQHRVQCRIVEHLTALVLRPIAETCOLSIE
NFQVYREPMENDTLCGFIIFSPSSCHRFSGKSVYIVFKGLRSIIAKCESSGQEAGL
LQDQLNLYACFTITDVAVSLSGEDRVLEIKELQRIYQVWGEFVQSGEASLEVRD
VRRLAVNRIEKAQHAELVRWDQYQAQCSMNPQVRLRAAI PHKNVFNENKLVNR
KYSRGEHLVRFGTDFVCGKQIRLAFRDHOGNLTVDKEGIDKVSDELVARLKGVSEV
GMOREGHIQVSPGANLSADILGTSKMSFHVNVNQFSRSPFLRYEVOTFDLYLWF
TASRLSKSPDNHNLATIFHENQGFIPANVRAAIKLEAGLNFSEKHEMSACLD
TQVSMIAVEGSGEENPLMIVFNHAGELASKNIRPEFAVGEVYLVNFGVDKAVS
LDSKETVQVQFHWTSKFCQGVNKTNSQFSQGRGWRMAVLDGYISDPVLNVLK
DHASVGNFYSREVRNLASDLKSGAMSFIPILSEEVVSPELGHSORLQILSVFLGL
VNLIALMSVYKFGVITASVAVLNLIALIWASMOYLDAPLTLISLAGIILAMGAVDA
NVLVFEERIREYLLSRSLSSVEGAYKAFSAIPDSNLTILASALLMLDTPGIFG
ALTLIGIFSMTLSLFTWTFELFIQKTRTQTLNMMKPIGVKHNFLKECKRLVWV
SGAVVVLGICLGGAWDSILGMDFGGYALTLDSDVCAYNPQMSALKKRFREIGL
SSRDYRVRKADSSGKVIYFSQNALARVEQAQKGSQNGADHHLAQVLQVLDSSDA
TSPAAPFSSRSGNFWKVSQGLSNKRTQAIWALFGALGIIILYVSLRPEWRYAFSAICA
LMHDLATCSVLVAFHFLHRIQIDLAQVAGLMTVLGYSLNNTLIIIFDRIDRREKL
FTWPIPLINDALQKTLQRTVMTATTLISLVILLFVGGSGIFNPAFIMTVGILLTUS
SLYIAPLLFLMWRKESQNSLQ"
BASE COUNT 3968 a 2882 g 3518 t
ORIGIN
Query Match 1.18; Score 20; DB 1; Length 12442;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 647 ATTAAGGGTTTGCTTGCAC 666
|||||
Db 9133 ATTAAGGGTTTGCTTGCAC 9114
|||||
RESULT 40
CEFS2H3/c 42632 bp DNA linear INV 24-JAN-2002
LOCUS
DEFINITION Caenorhabditis elegans cosmid F52H3, complete sequence.
ACCESSION Z66512
VERSION Z66512.1 GI:1041319
```

```
HTG.
Caenorhabditis elegans.
Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
1
none.
Genome sequence of the nematode C. elegans: a platform for
investigating biology. The C. elegans Sequencing Consortium
Science 282 (5396), 2012-2018 (1998)
95069613
9851916
The C.elegans Sequencing Consortium.
REFERENCE
2 (bases 1 to 42632)
Gardner,A.E.
Direct Submissions
Submitted (26-OCT-1995) Nematode Sequencing Project, Sanger
Institute, Hinxton, Cambridge CB10 1SA, England and Department of
Genetics, Washington University, St. Louis, MO 63110, USA. E-mail:
jes@sanger.ac.uk or rwnematode.wustl.edu
Coding sequences below are predicted from computer analysis, using
predictions from Genefinder (P. Green, U. Washington), and other
available information.
Current sequence finishing criteria for the C. elegans genome
sequencing consortium are that all bases are either sequenced
unambiguously on both strands, or on a single strand with both a
dye primer and dye terminator reaction, from distinct subclones.
Exceptions are indicated by an explicit note.
IMPORTANT: This sequence is not the entire insert of clone F52H3.
It may be shorter because we only sequence overlapping sections
once, or longer because we arrange for a small overlap between
neighbouring submissions.
The true left end of clone F52H3 is at 1 in this sequence. The true
right end of clone F52H3 is at 1981 in
sequence Z48543.
The true left end of clone C18D1 is at 42533 in this sequence. The
true right end of clone ZK892 is at 7399 in this sequence. The
start of this sequence (1..104) overlaps with the end of sequence
Z48638.
The end of this sequence (42533..42632) overlaps with the start of
sequence Z48543.
For a graphical representation of this sequence and its analysis
see: - http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?
name=F52H3
IMPORTANT: This sequence is NOT necessarily the entire insert of
the specified clone. It may be shorter because we only sequence
overlapping sections once, or longer because we arrange for a small
overlap between neighbouring submissions.
Location/Qualifiers
1..42632
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/chromosome="II"
/clone="F52H3"
join(Z48638.1:30587..30627,Z48638.1:30651..30750,
Z48638.1:30795..30954,Z48638.1:31010..31190,105..230,
324..405,449..601)
/gene="ZK892.5"
/codon_start=1
/protein_id="CAA91328.1"
/db_xref="GI:3877396"
/translacion="MELVSFRAVTHISCOLKEYPEDATEIRELIERRRHYLVNVSAFA
ILATFQWCTTVGVYIVLSAYAHMADSLVHISPMITTVFVLSVYPSIFPKKCAL
LVPLTLLYFAAASLKLQIANQHLDFGRHFKSTTKYSEYGLLSLFDYSILVFSI
INCAMCIGSWHFTPLVYVWNNFRKTEEARALRGLKTKKUSNCITTSNDSNAHK
LSHGAKGKRFISKHKQKTVQAQINENADVTTTSGAMDVKVGNQLFKLEIFWFOQE
VRVA"
complement(join(1113..1226,1296..1418,1464..1625,
```

CDS

1679.. 2006,2074.. 2162,2485.. 2553,4200.. 71031)
/gene="F52H3.7a"
complement(join(1113..1226,1296..1418,1464..1625,
1679..2006,2074..2162,2485..2553,4200..71031))
/note="contains similarity to Pfam domain: PF00337
(vertebrate galactoside-binding lectins), Score=231.0,
E-value=4.5e-68, N=2

CDNA EST yk6a5.5 comes from this gene
CDNA EST yk6f5.5 comes from this gene
CDNA EST yk25e11.5 comes from this gene
CDNA EST yk44h11.5 comes from this gene
CDNA EST yk46c9.5 comes from this gene
CDNA EST yk46f4.5 comes from this gene
CDNA EST EMBL:M89116 comes from this gene
CDNA EST EMBL:M89121 comes from this gene
CDNA EST yk6f5.3 comes from this gene
CDNA EST yk6a5.3 comes from this gene
CDNA EST yk44b10.3 comes from this gene
CDNA EST yk46c9.3 comes from this gene
CDNA EST yk83h1.5 comes from this gene
CDNA EST yk81g2.5 comes from this gene
CDNA EST yk11463.5 comes from this gene
CDNA EST EMBL:T00726 comes from this gene
CDNA EST EMBL:T02432 comes from this gene
CDNA EST yk25e11.3 comes from this gene
CDNA EST yk44h11.3 comes from this gene
CDNA EST yk46f4.3 comes from this gene
CDNA EST yk4f10.5 comes from this gene
CDNA EST yk122c8.5 comes from this gene
CDNA EST yk122c8.3 comes from this gene
CDNA EST yk14h8.5 comes from this gene
CDNA EST yk669g1.3 comes from this gene
CDNA EST yk497f10.5 comes from this gene"

/codon_start=1

/protein_id="CA91327.1"

/db_xref="GI:3877395"

/db_xref="SPTREMBL:O20684"

/translation="MADSSFFSSAFSSAKNLIPPAKNTEDKPVADTASDAKDADS
AIOEAVONASPAINDVFTDAVDKLEEDKDTVOFGNGDKKEEAAVEEKFDSVHYGV
DGETAEOEAEKEAQQVQVGLADGTQVIEGVTVAGTSDSLTNPSPITVSAP
VADEKKEETVAALDLSAPVBEKESHSSEHTAEASIDSIVQADPEARIVFOGL
ADGTQITIEASANPICOESAPAAVSEEVGAHVVPVDOAPEARTSADQADAE
EHEVAVHVNASTPRAVDHEETKAEERTPVVDSAPAOETIKVEHEKKEAADA
SAPAAHDSVHEETAAQVDEVEIAVHVASNESAPBEHKEVAAAATSEDKPRPAS
DEHTESAAGAISEVHIALTESAPASDNTTEAPAGVIEEKNEDVHDVDASSDA
PADHREAAASSESEVPIITAVEATESAAPADKAEETKIDSAATADADAVVAS
PADDOASEVTEAPHVESKASGAPATSEDKEDVQATSEVETOQAPADEHDDASPAEP
TTVQDVIETAPIYTESTPRAHESAPASBALAEHDEAAPSEPOVDVVEVETAP
HIAPEASAPASVEEAPSESTQAPVEHKAOTTTADSDSHEPRAESDAPAESEVETAP
VEHKEAEPALAESVEDVIEAVHVASNESAPAEVEDASADSTQAPVEEOKTAAAH
EPSEHQAADHAASVEDVIEAVHVASNESAPTEEAASESIQADHDEASTAEPAH
EESVTVASHKEPESAPVEENIESVAVSASEAPESGDAAPESAPVEENIEPAH
EVAVAVHVESITSEVPTDAGSEEVPOOQHTGAEEAASVDEESPDVHVDADAAP
ETATPESGSEETPEKTVHISOVEAPASEEHNSSEPELVEVDEVOIYAAVASAPETOP
EESPEKPRKLAEHMAHESAFVPRSKLVEPEPEPQTIIVYKGTGSDIRFTVNLHMT
ADPSGNDVPLHVSVRPDEGKLVENTFSEKSGEERSKPKKGDIDITIRADSKAF
QIFVQKEVEKEHVRPLSSITHTPTIGDVLVYVINGGKYYVPVYSGLAGEGLAG
KTLIVFGIPEKKAAPHNLKKNGLDILNARFDEKAVVRRSLINSAGNERBEK
MPPEKAVGFDEIHNEBPYAFVTVNGERFASVYARLSPDEVNGIQIGDVEITGICQA
"

gene

CDS

complement(join(1113..1226,1296..1418,1464..1625,
1679..2006,2074..2162,2485..25051)
/gene="F52H3.7b"
complement(join(1113..1226,1296..1418,1464..1625,
1679..2006,2074..2162,2485..25051)
/gene="F52H3.7b"
/note="similar to galactoside-binding lectin
CDNA EST EMBL:AB038503 comes from this gene"
/codon_start=1
/protein_id="CAC70097.1"
/db_xref="GI:15718128"
/translation="MAHRSAPVPRYRSKLVPEPFGQTLIVKGTGSDSIRFTVNLHMT

gene

CDS

NTADSGNDVPLHVSVRPDEGKLVENTFSGENGEKERSNPYKKGDDIDIRAPHS
KFOIFVQKEVEKEHVRPLSSITHTPTIGDVLVYVINGGKYYVPVYSGLAGEGLA
PGTLIVFGIPEKKAAPHNLKKNGLDILNARFDEKAVVRRSLINSAGNERBE
GMPPEKAVGFDEIHNEBPYAFVTVNGERFASVYARLSPDEVNGIQIGDVEITGIC
LA"
complement(join(17870..18124,18171..18458,18506..18784,
18863..18990,19044..190831))
/gene="F52H3.6"
complement(join(17870..18124,18171..18458,18506..18784,
18863..18990,19044..190831))
/gene="F52H3.6"
/note="contains similarity to Pfam domain: PF00149
(Ser/Thr protein phosphatase), Score=494.2,
E-value=3.3e-145, N=1"
/codon_start=1
/protein_id="CA91326.1"
/db_xref="GI:3877394"

gene

CDS

/translation="MVNIPSFIDRLSVNDTSNPSLRCINEADIDIVIECTKLLA
QATWEVQAPVAVCGDIGHQYTDLRIFNRCSPFDQVYLFGLVDVDRQOLEVTC
LMAYKIRYRPNRFTLRGNHCASINRTYGFYDECKRRSLALYNDPQNLPLCAM
ISGRIFCMHGGLSOKLVSWTQLAETTRFDPNPNSLAMDLMADPENMTGMAESSRG
VSQIFGADVVDPEPKMNDILARGHOVQDQGFEPFADKRLVTTIFSAKRCGEFNNNA
AMVLDERSLIVSFILKPAIREVKIQAGAAAKMGRSLTVSGRGSKXKG"
join(21426..21628,21681..21760,21812..21933,21978..22112)
/gene="F52H3.5"
join(21426..21628,21681..21760,21812..21933,21978..22112)
/gene="F52H3.5"
/note="contains similarity to Pfam domain: PF00515 (TPR
Domain), Score=60.9, E-value=9.1e-15, N=3
CDNA EST yk573e3.5 comes from this gene"
/codon_start=1
/protein_id="CA91325.1"
/db_xref="GI:3877393"
/db_xref="SPTREMBL:O20683"
/translation="WSTSHDRVYLNQIINPLMPTSDSGIAEIHLESDEHPGVELSL
QLERGVALLAEGRVLDLAEIKETKALECPKPSVYNNKQAYTRQNKREKALDLNE
ALSLAGPKTKTACQAVYORASIVYRLRGDDKARTFPASAAELGSSFAKQVLALNPYA
AMCNMLAEVFEKATGED"
complement(join(22244..22462,22519..22958,23022..23425,
23475..23858,23912..24182,24230..244281))
/gene="F52H3.2"
complement(join(22244..22462,22519..22958,23022..23425,
23475..23858,23912..24182,24230..244281))
/gene="F52H3.2"
/note="contains similarity to Pfam domain: PF01134
(glucose inhibited division protein A), Score=1405.3,
E-value=0, N=1
CDNA EST yk498a2.5 comes from this gene"
/codon_start=1

gene

CDS

Query Match 1.1%; Score 20; DB 3; Length 42632;
Best Local Similarity 100.0%; Pred No. 38; 0; Indels 0; Gaps 0;
Matches 20; Conservative 0; Mismatches 0;
Oy 1397 TTCTTTTGAAGAAATTC A 1416
Db 7940 TTCTTTTGAAGAAATTC A 7921

RESULT 41

AC119990/c 72371 bp DNA linear HTG 02-MAY-2002

DEFINITION Mus musculus clone RP24-506H15, low-PASS SEQUENCE SAMPLING.

AC119990

VERSION AC119990.1 GI:20389650

KEYWORDS HTG; HTGS PHASE0.

SOURCE

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 72371)

AUTHORS Birren,B., Linton,L., Nuebaum,C. and Lander,E.

TITLE JOURNAL REFERENCE AUTHORS

Mus musculus, clone RP24-506H15
Unpublished
2. (bases 1 to 72371)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavsky,L.,
Boukhalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collamore,A.,
Cook,A., Cooke,P., DeArrellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kelle,C., LaRocque,K., Lamazares,R.,
Landers,T., Lehoczy,J., Levine,R., Lindblad-Toh,K., Liu,G.,
MacLean,C., MacDonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Roberti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Srauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (02-MAY-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Snit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

TITLE JOURNAL COMMENT

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L25778
Center clone name: 506_H_15

* NOTE: This record contains 88 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

1 723: contig of 723 bp in length
724 823: gap of 100 bp
824 1555: contig of 732 bp in length
1556 1655: gap of 100 bp
1656 2380: contig of 725 bp in length
2381 2480: gap of 100 bp
2481 3206: contig of 726 bp in length
3207 3306: gap of 100 bp
3307 4044: contig of 738 bp in length
4045 4144: gap of 100 bp
4145 4861: contig of 717 bp in length
4862 4961: gap of 100 bp
4962 5695: contig of 734 bp in length
5696 5795: gap of 100 bp
5796 6538: contig of 743 bp in length
6539 6638: gap of 100 bp
6639 7370: contig of 732 bp in length
7371 7470: gap of 100 bp
7471 8197: contig of 727 bp in length
8198 8297: gap of 100 bp
8298 9028: contig of 731 bp in length
9029 9128: gap of 100 bp

9129 9864: contig of 736 bp in length
9865 9964: gap of 100 bp
9965 10685: contig of 701 bp in length
10686 10765: gap of 100 bp
10766 11493: contig of 728 bp in length
11494 11593: gap of 100 bp
11594 12309: contig of 716 bp in length
12310 12409: gap of 100 bp
12410 13151: contig of 742 bp in length
13152 13251: gap of 100 bp
13252 13983: contig of 732 bp in length
13984 14083: gap of 100 bp
14084 14810: contig of 727 bp in length
14811 14910: gap of 100 bp
14911 15642: contig of 732 bp in length
15643 15742: gap of 100 bp
15743 16403: contig of 661 bp in length
16404 16503: gap of 100 bp
16504 17225: contig of 722 bp in length
17226 17325: gap of 100 bp
17326 18042: contig of 717 bp in length
18043 18142: gap of 100 bp
18143 18858: contig of 716 bp in length
18859 18958: gap of 100 bp
18959 19778: contig of 720 bp in length
19779 19778: gap of 100 bp
19779 20512: contig of 734 bp in length
20513 20612: gap of 100 bp
20613 21349: contig of 737 bp in length
21350 21449: gap of 100 bp
21450 22187: contig of 738 bp in length
22188 22287: gap of 100 bp
22288 23012: contig of 725 bp in length
23013 23112: gap of 100 bp
23113 23851: contig of 739 bp in length
23852 23951: gap of 100 bp
23952 24685: contig of 734 bp in length
24686 24785: gap of 100 bp
24786 25514: contig of 729 bp in length
25515 25614: gap of 100 bp
25615 26336: contig of 722 bp in length
26337 26436: gap of 100 bp
26437 27155: contig of 719 bp in length
27156 27255: gap of 100 bp
27256 27991: contig of 736 bp in length
27992 28091: gap of 100 bp
28092 28816: contig of 725 bp in length
28817 28916: gap of 100 bp
28917 29638: contig of 722 bp in length
29639 29738: gap of 100 bp
29739 30456: contig of 718 bp in length
30457 30556: gap of 100 bp
30557 31304: contig of 748 bp in length
31305 31404: gap of 100 bp
31405 32123: contig of 719 bp in length
32124 32223: gap of 100 bp
32224 32954: contig of 731 bp in length
32955 33054: gap of 100 bp
33055 33785: contig of 731 bp in length
33786 33885: gap of 100 bp
33886 34566: contig of 681 bp in length
34567 34666: gap of 100 bp
34667 35384: contig of 718 bp in length
35385 35484: gap of 100 bp
35485 36200: contig of 716 bp in length
36201 36300: gap of 100 bp
36301 37019: contig of 719 bp in length
37020 37119: gap of 100 bp
37120 37848: contig of 729 bp in length
37849 37948: gap of 100 bp
37949 38670: contig of 722 bp in length
38671 38770: gap of 100 bp
38771 39505: contig of 735 bp in length


```

misc_feature /note="assembly_fragment"
1225..2355
misc_feature /note="assembly_fragment"
2456..5644
misc_feature /note="assembly_fragment"
5745..16114
misc_feature /note="assembly_fragment"
16215..23188
misc_feature /note="assembly_fragment"
23289..73256
BASE COUNT 23533 a 13938 c 13168 g 22117 t 500 others
ORIGIN
Query Match 1.1%; Score 20; DB 2; Length 73256;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1395 ATTCTTTTGAAGAAATT 1414
|||||
Db 17872 ATTCTTTTGAAGAAATT 17891

RESULT 43
AC087648/c 74105 bp DNA linear HTG 15-JAN-2001
LOCUS Homo sapiens chromosome 17 clone RP11-21M3 map 17, LOW-PASS
SEQUENCE SAMPLING.
ACCESSION AC087648
VERSION 1 GI:12229423
KEYWORDS HTG; HTGS PHASE0.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 74105)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 17, clone RP11-21M3
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 74105)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
Barna,N., Bahtien,V., Boguslavsky,L., Bouckgalter,B., Brown,A.,
Camarata,J., Campopiano,A., Choepel,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., DeArlano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J.,
Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Karatas,A., Labrecque,K., Lamazares,R., Landers,T.,
Lehoczky,J., Levine,R., Liu,G., MacLean,C., Macdonald,P.,
Marquis,N., Matthews,C., McCarthy,M., McSwan,P., McKernan,K.,
McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R.,
Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M.,
Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
Sougnuez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A.,
Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (15-JAN-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT All repeats were identified using RepeatMasker:
Snit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information

```

Center project name: L11999
Center clone name: 21_M_3

* NOTE: This record contains 88 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

1 731: contig of 731 bp in length
* 732 831: gap of 100 bp
* 832 1595: contig of 754 bp in length
* 1596 1685: gap of 100 bp
* 1686 2373: contig of 688 bp in length
* 2374 2473: gap of 100 bp
* 2474 3236: contig of 763 bp in length
* 3237 3336: gap of 100 bp
* 3337 4056: contig of 720 bp in length
* 4057 4156: gap of 100 bp
* 4157 4913: contig of 757 bp in length
* 4914 5013: gap of 100 bp
* 5014 5752: contig of 739 bp in length
* 5753 5852: gap of 100 bp
* 5853 6597: contig of 745 bp in length
* 6598 6697: gap of 100 bp
* 6698 7443: contig of 746 bp in length
* 7444 7543: gap of 100 bp
* 7544 8308: contig of 765 bp in length
* 8309 8408: gap of 100 bp
* 8409 9189: contig of 781 bp in length
* 9190 9289: gap of 100 bp
* 9290 10040: contig of 751 bp in length
* 10041 10140: gap of 100 bp
* 10141 10887: contig of 747 bp in length
* 10888 10987: gap of 100 bp
* 10988 11730: contig of 743 bp in length
* 11731 11830: gap of 100 bp
* 11831 12550: contig of 720 bp in length
* 12551 12650: gap of 100 bp
* 12651 13388: contig of 738 bp in length
* 13389 13488: gap of 100 bp
* 13489 14223: contig of 735 bp in length
* 14224 14323: gap of 100 bp
* 14324 15061: contig of 738 bp in length
* 15062 15161: gap of 100 bp
* 15162 15898: contig of 737 bp in length
* 15899 15998: gap of 100 bp
* 15999 16748: contig of 750 bp in length
* 16749 16848: gap of 100 bp
* 16849 17573: contig of 725 bp in length
* 17574 17673: gap of 100 bp
* 17674 18410: contig of 737 bp in length
* 18411 18510: gap of 100 bp
* 18511 19260: contig of 750 bp in length
* 19261 19360: gap of 100 bp
* 19361 20103: contig of 743 bp in length
* 20104 20203: gap of 100 bp
* 20204 20970: contig of 767 bp in length
* 20971 21070: gap of 100 bp
* 21071 21819: contig of 749 bp in length
* 21820 21919: gap of 100 bp
* 21920 22662: contig of 743 bp in length
* 22663 22762: gap of 100 bp
* 22763 23477: contig of 715 bp in length
* 23478 23577: gap of 100 bp
* 23578 24318: contig of 741 bp in length
* 24319 24418: gap of 100 bp
* 24419 25168: contig of 750 bp in length

```

* 25169 25268: gap of 100 bp
* 25269 26011: contig of 743 bp in length
* 26012 26111: gap of 100 bp
* 26112 26864: contig of 753 bp in length
* 26865 26964: gap of 100 bp
* 26965 27708: contig of 744 bp in length
* 27709 27808: gap of 100 bp
* 27809 28557: contig of 749 bp in length
* 28558 28657: gap of 100 bp
* 28658 29405: contig of 748 bp in length
* 29406 29505: gap of 100 bp
* 29506 30249: contig of 744 bp in length
* 30250 30349: gap of 100 bp
* 30350 31093: contig of 744 bp in length
* 31094 31193: gap of 100 bp
* 31194 31937: contig of 744 bp in length
* 31938 32037: gap of 100 bp
* 32038 32710: contig of 733 bp in length
* 32711 32870: gap of 100 bp
* 32871 33677: contig of 757 bp in length
* 33678 33727: gap of 100 bp
* 33728 34471: contig of 744 bp in length
* 34472 34571: gap of 100 bp
* 34572 35288: contig of 727 bp in length
* 35289 35398: gap of 100 bp
* 35399 36145: contig of 747 bp in length
* 36146 36245: gap of 100 bp
* 36246 36980: contig of 735 bp in length
* 36981 37080: gap of 100 bp
* 37081 37818: contig of 738 bp in length
* 37819 37918: gap of 100 bp
* 37919 38663: contig of 745 bp in length
* 38664 38763: gap of 100 bp
* 38764 39504: contig of 741 bp in length
* 39505 39604: gap of 100 bp
* 39605 40351: contig of 747 bp in length
* 40352 40451: gap of 100 bp
* 40452 41172: contig of 721 bp in length
* 41173 41272: gap of 100 bp
* 41273 42013: contig of 741 bp in length
* 42014 42113: gap of 100 bp
* 42114 42857: contig of 744 bp in length
* 42858 42957: gap of 100 bp
* 42958 43634: contig of 727 bp in length
* 43635 43784: gap of 100 bp
* 43785 44527: contig of 743 bp in length
* 44528 44627: gap of 100 bp
* 44628 45356: contig of 739 bp in length
* 45357 45466: gap of 100 bp
* 45467 46216: contig of 750 bp in length
* 46217 46316: gap of 100 bp
* 46317 47015: contig of 699 bp in length
* 47016 47115: gap of 100 bp
* 47116 47823: contig of 708 bp in length
* 47824 47923: gap of 100 bp
* 47924 48666: contig of 743 bp in length
* 48667 48766: gap of 100 bp
* 48767 49506: contig of 740 bp in length
* 49507 49606: gap of 100 bp
* 49607 50350: contig of 744 bp in length
* 50351 50450: gap of 100 bp
* 50451 51179: contig of 729 bp in length
* 51180 51279: gap of 100 bp
* 51280 52021: contig of 742 bp in length
* 52022 52121: gap of 100 bp
* 52122 52850: contig of 729 bp in length
* 52851 52950: gap of 100 bp
* 52951 53684: contig of 734 bp in length
* 53685 53784: gap of 100 bp
* 53785 54535: contig of 751 bp in length
* 54536 54635: gap of 100 bp
* 54636 55392: contig of 757 bp in length
* 55393 55492: gap of 100 bp

```

```

* 55493 56262: contig of 770 bp in length
* 56263 56362: gap of 100 bp
* 56363 57105: contig of 743 bp in length
* 57106 57205: gap of 100 bp
* 57206 57966: contig of 761 bp in length
* 57967 58066: gap of 100 bp

Query Match 1.1%; Score 20; DB 2; Length 74105;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1410 AATTCGAATGATTGGCA 1429
Db 48377 AATTCGAATGATTGGCA 48358

RESULT 44
AL713969
LOCUS Human DNA sequence from clone RP11-501117 on chromosome 10,
DEFINITION complete sequence.
ACCESSION AL713969
VERSION AL713969
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
1 (bases 1 to 84574)
Heath,P.
REFERENCE Direct Submission
AUTHORS Submitted (31-MAY-2002) Wellcome Trust Sanger Institute, Hinxton,
TITLE Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
JOURNAL humer@esanger.ac.uk Clone requests: clonerequests@sanger.ac.uk
On Jun 21, 2002 this sequence version replaced gi:21212454.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em: EMBL; Sw:
SWISSPROT; Tr: TrEMBL; Wp: WormPEP; information on the WORMPEP
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormep
This sequence was generated from part of bacterial clone contigs of human
chromosome 10, constructed by the Sanger Centre Chromosome 10
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr10
RP11-501117 is from the library RPI1-11.2 constructed by the group
of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBACe3.6.

FEATURES
source
location/Qualifiers
1..84574
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/clone="RP11-501117"
/clone_lib="RPI1-11.2"
BASE COUNT 27476 a 16456 c 15040 g 25602 t
ORIGIN
Query Match 1.1%; Score 20; DB 9; Length 84574;
Best Local Similarity 100.0%; Pred. No. 38;

```


GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: December 17, 2002, 12:32:52 ; Search time 17.5 Seconds
(without alignments)
6267.923 Million cell updates/sec

Title: US-09-868-987-1
Perfect score: 3312
Sequence: 1 atggaattccgattatgtc.....atgacaaattcagataatgc 1864

Scoring table: BLOSUM62

Xgapop 10.0	Xgapext 0.5
Xgapop 10.0	Xgapext 0.5
Xgapop 6.0	Xgapext 7.0
Deiop 6.0	Delext 7.0

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 525148

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlh
-Q=/cgcn2_1/USPTO.spool/US09868987/runat 16122002 083007 23590/app query.fasta_1.2055
-DB=Issued Patents AA -OPMT=fastan -SUFFIX=rai -MINMATCH=0.1 -LOEPCI=0
-LOEPCXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -LOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=45
-MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09868987 @CGN 1 1 0 @runat 16122002 083007 23590 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGESUBRY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WATN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FCGAPOP=6 -FCGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued Patents AA: *
1: /cgcn2_6/ptodata/1/iaa/5A.COMB.pcp: *
2: /cgcn2_6/ptodata/1/iaa/5B.COMB.pcp: *
3: /cgcn2_6/ptodata/1/iaa/6A.COMB.pcp: *
4: /cgcn2_6/ptodata/1/iaa/6B.COMB.pcp: *
5: /cgcn2_6/ptodata/1/iaa/PTCUS.COMB.pcp: *
6: /cgcn2_6/ptodata/1/iaa/backfiles1.pcp: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	678.5	20.5	737	4	US-09-462-844-2
2	647.5	19.6	766	4	US-09-134-001C-3689
3	364	11.0	323	4	US-09-462-844-4
4	359.5	10.9	615	4	US-09-462-844-3
5	136.5	4.1	1051	4	US-09-134-001C-5005
6	117	3.5	474	3	US-09-022-699-2
7	108	3.3	901	4	US-09-134-001C-5389
8	103.5	3.1	383	2	US-08-196-989B-4
9	103.5	3.1	383	2	US-08-760-936-4
10	101	3.0	381	2	US-08-845-566-3
11	101	3.0	381	2	US-08-467-948A-28
12	101	3.0	381	3	US-08-852-824-18

13	101	3.0	381	3	US-08-467-947A-28	Sequence 28, Appl
14	101	3.0	381	5	PCT-US96-10618-4	Sequence 4, Appl
15	99.5	3.0	524	4	US-09-134-001C-3353	Sequence 3353, Ap
16	97.5	2.9	382	4	US-09-542-733-2	Sequence 2, Appl
17	97.5	2.9	1820	3	US-07-998-289B-8	Sequence 8, Appl
18	97.5	2.9	2100	2	US-08-808-793-23	Sequence 23, Appl
19	97.5	2.9	2100	3	US-08-772-512A-19	Sequence 19, Appl
20	97	2.9	400	4	US-09-134-001C-2912	Sequence 2912, Ap
21	97	2.9	1090	3	US-08-307-896-3	Sequence 3, Appl
22	97	2.9	1090	3	US-08-726-214-4	Sequence 4, Appl
23	97	2.9	1090	5	PCT-US95-11808-3	Sequence 3, Appl
24	96.5	2.9	382	4	US-09-262-477-2	Sequence 2, Appl
25	95	2.9	518	4	US-09-134-001C-4744	Sequence 4744, Ap
26	94	2.8	352	1	US-08-196-989B-2	Sequence 2, Appl
27	94	2.8	352	2	US-08-760-936-2	Sequence 2, Appl
28	93.5	2.8	490	4	US-09-134-001C-5116	Sequence 5116, Ap
29	93.5	2.8	1285	4	US-08-540-406-6	Sequence 6, Appl
30	93.5	2.8	1285	3	US-08-656-055-6	Sequence 6, Appl
31	93.5	2.8	1285	4	US-08-954-668-6	Sequence 6, Appl
32	93.5	2.8	1285	4	US-08-918-658-6	Sequence 6, Appl
33	93.5	2.8	1285	5	PCT-US95-13233-6	Sequence 6, Appl
34	93.5	2.8	1286	4	US-09-268-140-3	Sequence 3, Appl
35	93.5	2.8	1299	4	US-08-460-900C-62	Sequence 62, Appl
36	93.5	2.8	1299	4	US-08-674-509B-48	Sequence 48, Appl
37	93.5	2.8	1299	4	US-08-954-698-48	Sequence 48, Appl
38	93	2.8	344	4	US-09-134-001C-5567	Sequence 5567, Ap
39	92.5	2.8	334	1	US-08-118-270-73	Sequence 73, Appl
40	92.5	2.8	334	5	PCT-US93-08528-73	Sequence 73, Appl
41	92	2.8	335	4	US-09-387-699-2	Sequence 2, Appl
42	92	2.8	335	4	US-09-641-259B-2	Sequence 2, Appl
43	91	2.7	269	1	US-08-118-270-64	Sequence 64, Appl
44	91	2.7	269	5	PCT-US93-08528-64	Sequence 64, Appl
45	91	2.7	341	2	US-08-846-762-92	Sequence 92, Appl

ALIGNMENTS

RESULT 1
US-09-462-844-2
; Sequence 2, Application US/09462844
; Patent No. 6258563
; GENERAL INFORMATION:
; APPLICANT: Quax, Wilhelmus J.
; TITLE OF INVENTION: Increasing Production of Proteins in Gram-Positive Microorganisms
; FILE REFERENCE: GC385-US
; CURRENT APPLICATION NUMBER: US/09/462,844
; CURRENT FILING DATE: 2000-01-13
; PRIOR APPLICATION NUMBER: PCT/US98/14786
; PRIOR FILING DATE: 1998-07-15
; PRIOR APPLICATION NUMBER: EP 97305286.3
; PRIOR FILING DATE: 1997-07-16
; PRIOR APPLICATION NUMBER: EP 97305344.0
; PRIOR FILING DATE: 1997-07-17
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 737
; TYPE: PRT
; ORGANISM: Bacillus subtilis
US-09-462-844-2

Alignment Scores:
Pred. No.: 3.82e-69 Length: 737
Score: 678.50 Matches: 169
Percent Similarity: 50.43% Conservative: 125
Best Local Similarity: 28.99% Mismatches: 192
Query Match: 20.49% Indels: 97
DB: 4 Gaps: 15

US-09-868-987-1 (1-1864) x US-09-462-844-2 (1-737)
Qy 104 GTCAGCAGCCCTATTTTAAACGTCCTGAGGAAA ---AATCATGCCAGTGTCTCAGGAAA 160

```

Db 197 ValSerIaProaenValSerIngluLeuAsnThrThrAspValysIleIleGluHis 216
QY 161 TTTACCAACCGTGAAGTGAAGCAAACTGCGCTCAGATTAAATGTGAGCGATCTTTT 220
Db 217 PheThrIaIeIngluIaIaIaYAspLeuAlaSerIleLeuAsnIaIaIaLeu----- 234
QY 221 GTTCCCGAGGTTCTCAGTGAAGAG-----ACGATCTCTTCATCTTTGGAAAAA 271
Db 235 ---ProValysLeuThrIeugIuIySerThrSerValIaIaIaIaIaIaIaIaIa 253
QY 272 CAATGTACACAAAGCATTATCTCAGACATGCTGAGCTTGAGCAATGCTTATGTTGATG 331
Db 254 AlaLeuHisAspThrValPheAlaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 273
QY 332 AGCGTATATATATAGATTGTGAGGCGTCACTCGGAGCTGTCTTGTGAATCTTTTG 391
Db 274 LeuPheThrIyTrIyArgLeuPProGlyLeuIle-----AlaValIleThrLeuSerVal 290
QY 392 CTTATCTGGCAGCTCTACAG-----TATTTGATGGCGCACTCCTTGTACAGA 442
Db 291 TyrIleThrIeThrLeuGlnIlePheAspTrpMetAsnIaIaValLeuThrLeuProGly 310
QY 443 CTGCGTGGAGATTGTTCTGTATGAGGATGCGGATGCGATGCAATGCTTGTATTCGAA 502
Db 311 IleAlaIaLeuIleLeuGlyValaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 330
QY 503 AGATCCGAGAGAAATTTTATTTCTCAAGTCTTAAATCTTGAAGAAAGATAT 562
Db 331 ArgIleYsIeGluIuLeuIyLeuGlyLysSerValaArgSerAlaPheArgSerGlyAsn 350
QY 563 ACCAAGCTTTTGAGGCACTTTTGTATCTTACCTTGACTACAGTATGGCTCAGCACT 622
Db 351 ArgArgSerPheAlaThrIlePheAspAlaAsnIleThrIleIleAlaIaIaVala 370
QY 623 CTTTCTTCCAGATACAG36CTATTAAGGTTTGCTTGACATGATTTTGAAT 682
Db 371 LeuIleIlePheGlyThrSerSerValIyLeuIaIaIaIaIaIaIaIaIaIaIaIa 390
QY 683 TTCTCTCAATGTTTACGCGTCTTTTCACTGACTAAATTTTCTTCATGCTG----- 733
Db 391 LeuThrSerPheIleThrIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 410
QY 734 -----TGATGAATAAG----- 745
Db 411 SerArgTrpLeuAspArgLysGlyTrpPheGlyValAsnLysLysIleMetAsp 430
QY 746 -----ACCAACATACACAGTTCATGATGATGATGATGATGATGATGATGATGAT 793
Db 431 IleGlnAspThrAspGluAsnThrGluProHisThrProPheGln-----Lys 446
QY 794 CATGATTTCTTGAGAGATGCAAAAACTTGGCGTGTCTTGAGAGTGTCTTTCTTTT 853
Db 447 TrpAspThrSerIyAspGlyIyPhePheIlePheSerSerAlaValaIaIa 466
QY 854 GGTTCGCTGCTCTCGGCTTTGAGCCTCGAATTCGTTTGGGAATGATTTAAAGGA 913
Db 467 GlyIleIleIleLeuValaPheArgLeuAsn-----LeuGlyIleAspPheAlaSer 484
QY 914 GGG-----TATGCTTTTACCTTATTCACAAAAGACATGCGC 949
Db 485 GlyAlaIaArgIleGluValaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 504
QY 950 ATCAGCATGTTGCTCAATTCGTCGCAAGTGTGCAATAACTACAGAGAGCTGTCTT 1009
Db 505 PheIuSerIeugIyMetAspProAspThrValaIaIaIaIaIaIaIaIaIaIaIaIa 524
QY 1010 TCT--TCTAGAGACTTCGATTCAACATTTGATCTTCAGAAAGATCAAAATCTAT 1066
Db 525 GlyValaIaArgPheValaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 544
QY 1067 TTTAGTATAAGCTTTAAGCTATACTAAGACATACAGAGCTCTCTCTTAAATTAACG 1126

```

```

Db 545 PheIyAspLys----- 548
QY 1127 ATCATGAGCTGGCTTATTTGTGGATTTGTCAGAAACAGGCTTGATTTCTTCGGA 1186
Db 549 ----- 550
QY 1187 AACCTTAAACGAAAGCAAAATTTTGTCAAGGTACAGAGCAAACTATACGAAGAAAG 1246
Db 551 SerAspProAsnValSerThrValaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 570
QY 1247 CGTTATCAGCGCAACATCGGCTTTTATGAGCTTTGGCAATCTTCTGTATGTAGT 1306
Db 571 LeuThrAlaValaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 586
QY 1307 TTGCGCTTGAATGCAAAATGCTTTTCAAGTCCGCTATGCGCTTAAATTCATGACCTTTG 1366
Db 587 IleArgPheGluIyTrpYsMetAlaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 604
QY 1367 GCTACCTGCGACGCTCTGTTTATAGCAATTTCTTTTGAAGAAATTCAAATAGATTG 1426
Db 605 -----AlaPhePheIleValaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 621
QY 1427 CAAGCATTTGCTTAAATGACTGATTTGGGATTCATTAACATATCTTGATCAT 1486
Db 622 ThrPheIleAlaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 641
QY 1487 TTGATGCTATTTGCTGAA-----GATCCGAAGCGAAGCTTTTACCTTATGAT 1537
Db 642 PheAspArgValaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 661
QY 1538 GTTTTGTATTAATGATGCGCTTCAAAAGACGTTACGCGGACGCTTATGCAACAGCTAC 1597
Db 662 HisIleValaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 681
QY 1598 ACTCTATCAAGTTTGTAAATGCTTTTGTATAGCGGCTCTCTCTCTTAAATTTTGA 1657
Db 682 ValValaIleValaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 701
QY 1658 TTTATATGACCATACAGATTTCTTACGACATTTATGCTCTTTTATATGACACACT 1717
Db 702 IleAlaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 721
QY 1718 CTGTGTG 1726
Db 722 IleTrpLeu 724

RESULT 2
US-09-134-001C-3689
; Sequence 3689, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCU
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3689
; LENGTH: 766
; TYPE: PRF
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3689

Alignment Scores:
Pred. No.: 1,45e-65 Length: 766
Score: 647.50 Matches: 165
Percent Similarity: 50.95% Conservative: 129
Best Local Similarity: 28.60% Mismatches: 204

```


Alignment Scores:

Pred. No.: 4.15e-33 Length: 323
 Score: 364.00 Matches: 102
 Percent Similarity: 49.52% Conservative: 53
 Best Local Similarity: 32.59% Mismatches: 124
 Query Match: 10.99% Indels: 34
 DB: 4 Gaps: 9

US-09-868-987-1 (1-1864) x US-09-462-844-4 (1-323)

QY 794 CATGATTTCTGAGAGATGCAAAAACCTTGGCGTGTCTTCTGGAAGTCTTTCTTTA 853
 Db 17 TyrAspPheMetArgTrpAspTyrTrpAlaPheGlyIleSerGlyLeuLeuLeuIleAla 36
 QY 854 GGTTCGCTGCTCGGGTGTGAGCTGGAATTCCTTTGGAAATGAGATTTAAAGGA 913
 Db 37 AlaIleValIleMetGlyValArgGlyPheAsn-----TrpGlyLeuAspPheThrGly 54
 QY 914 GGGTATGCTTACCTTAAATCCAAAGAGATGGCAGCAGCATGGTGTCTCAATGGCT 973
 Db 55 GlyThrValIleGluIleThrLeuGluIleProAlaGluIleAspVal-----MetArg 72
 QY 974 GGCAAGTTGTGATTAACACAGAGCTGTCTTCTTACAGAACTTCCGATTCAA 1033
 Db 73 -----AspAlaLeuGluIleValAspIlePhe-----GluGluProMetLeuGln 86
 QY 1034 ACATTTGGATCTTCAAGAAAGATCAAAATCTATTAGTAAAGCTTAAAGCTAACT 1093
 Db 87 AsnPheGlySerSerHisAspIleMetValArgMetProProAlaGluGlyIleThrGly 106
 QY 1094 AAGCAGATACAGACCTCT---CTCTAAATTAACGATCAGAGCTGGCGCTTATTTGGG 1150
 Db 107 GlyIleValIleGluIleSerGlnValLeuIleValIleAsnGluSer----- 121
 QY 1151 ATGTGTTCAGAAACAGCGCTAGATTTCTTACGGAACCTTAAACGAAACGAAATTT 1210
 Db 122 -----ThrAsnGlnAsnAlaIleValIleArg 130
 QY 1211 TGGTCAAGGTAAAGAGCAAACTATCGAAGAAATGCGTTATCGAGCAGCATGGGCTT 1270
 Db 131 IleGluPheValGlyProSerValGlyIleAspLeuAlaGlnThrGlyAlaMetAlaLeu 150
 QY 1271 TTAGAGCTTTGGCAATATCTGCTCTATGTGAGTTGGCTTGGATGGCAATATGCT 1330
 Db 151 MetAlaIleLeuSerIleLeuValIleValGlyPheArgPheGluTrpArgLeuAla 170
 QY 1331 TTCAGTCCGATGCGCTTAAATTCATGACCTTTGGCTTGGCTGCGAGCTTGTATTATA 1390
 Db 171 AlaGlyValIleIleLeuAlaHisAspValIleIleThrLeuGlyIleLeuSerIleu 190
 QY 1391 GCACATTTCTTTTGAAAGAAATTCAAATAGATTGCAAGCCATTGGCTTTAATGACT 1450
 Db 191 PheHis-----IleGluIleAspLeuThrIleValAlaSerIleuMetSer 205
 QY 1451 GTATTGGGGTATCTTAACAATACTTTGATCATTTTGTATCGGATTCGGAGAT--- 1507
 Db 206 ValIleGlyIleSerLeuAsnAspSerIleValIleSerAspArgIleArgGluAsnHe 225
 QY 1508 CGCAGAGGAAACCTGTTAAACCCCTATGATGATGTTAATGATGAGCCCTTCAAAAGAG 1567
 Db 226 ArgIleValIleArgArgGlyThrProIleGluIlePhe---AsnValSerLeuThrGlnThr 244
 QY 1568 TTCAGCCGACGATTAACAACAGCTCAATCTATACAGTTTGTATTGTTGTTT 1627
 Db 245 LeuHisArgThrIleIleThrSerGlyThrThrLeuMetValIleLeuMetLeuIleu 264
 QY 1628 ATAGCGGCTCTCTGCTCTTAATTTGATTTATTAAGACATAGGAGATCTTTAGGA 1687
 Db 265 PheGlyGlyProValIleGluGlyPheSerLeuThrMetLeuIleGlyValSerIleGly 284
 QY 1688 ACTTATGCTCTCTTATATGACACAGCTCTGTGTG 1726
 Db 285 ThrIleSerSerIleIleValIleSerAlaLeuAlaLeu 297

RESULT 4

US-09-462-844-3

/ Sequence 3, Application US/09462844
 / Patent No. 6258563
 / GENERAL INFORMATION:
 / APPLICANT: Quax, Wilhelmus J.
 / TITLE OF INVENTION: Increasing Production of Proteins in
 / Gram-Positive Microorganisms
 / FILE REFERENCE: GC385-US
 / CURRENT APPLICATION NUMBER: US/09/462,844
 / CURRENT FILING DATE: 2000-01-13
 / PRIOR APPLICATION NUMBER: PCT/US98/14786
 / PRIOR FILING DATE: 1998-07-15
 / PRIOR APPLICATION NUMBER: EP 97305286.3
 / PRIOR FILING DATE: 1997-07-16
 / PRIOR APPLICATION NUMBER: EP 97305344.0
 / PRIOR FILING DATE: 1997-07-17
 / NUMBER OF SEQ ID NOS: 11
 / SOFTWARE: FastSeq for Windows Version 4.0
 / SEQ ID NO 3
 / LENGTH: 615
 / TYPE: PR1
 / ORGANISM: Escherichia coli
 / US-09-462-844-3

Alignment Scores:

Pred. No.: 1.92e-32 Length: 615
 Score: 359.50 Matches: 80
 Percent Similarity: 62.39% Conservative: 61
 Best Local Similarity: 35.40% Mismatches: 82
 Query Match: 10.85% Indels: 3
 DB: 4 Gaps: 2

US-09-868-987-1 (1-1864) x US-09-462-844-3 (1-615)

QY 44 AATGACATATTTCTGAAACCGTGTGATGGCTGTAGATGATGACGGTATATG 103
 Db 375 SerGlyIleValIleAspAlaAsn-----GlyArgAlaValLeuValIleGlnGluVal 392
 QY 104 GTCAGAGCCCTATTTTAAACGTCCATTTGAAATTCATGCCAGTCTTCAAGGAAATTT 163
 Db 393 IleAsnIleAlaAsnIleGlnSerArgLeuGlyAsnSerPheArgIleThrGlyIleAsn 412
 QY 164 ACCCAGCTGAAGTGAACAAACTCGCTCGATTTAAATCTGAGCGAGATCTTTGTT 223
 Db 413 AsnProAsnGluIleIleArgGlnLeuSerLeuLeuIleValIleGlyAlaLeuIleAlaPro 432
 QY 224 CCGAGGTTCTCAGTGAAGAGAGATCTTCTGATCTTGGAAAAACAATGTACACAA 283
 Db 433 IleGlnIleValIleGluIleGluArgThrIleGlyProThrLeuGlyMetGlnAsnIleGluGln 452
 QY 284 GGCATTTCTCAGCATGCTGTGCGCTTGGCAATGCTTATTTGATGAGCGTATATAT 343
 Db 453 GlyLeuGluIleAlaCysLeuAlaGlyLeuLeuValSerIleLeuPheMetIleIlePheTyr 472
 QY 344 AGATTGGAGCGCTATCGCTTGGAGAGCTTCTTGTGAATCTTTGCTTATCTGGGCA 403
 Db 473 LysIlePheGlyLeuIleAlaThrSerAlaLeuIleAlaAsnLeuIleLeuIleValGly 492
 QY 404 GCTGTACAGTATTGG---GATGGCGCACCTCAGCTGTGACAGACCTCGGTGATTTGTTCT 460
 Db 493 IleMetSerLeuLeuProGlyAlaThrIleLeuSerMetProGlyIleAlaGlyIleValLeu 512
 QY 461 GCTATGGGATGCGCGTGAATGCAAAATGTTCTGTATTCGAAAGATCCAGAGGAATTT 520
 Db 513 ThrIleuAlaValAlaValAspAlaAsnValLeuIleAsnGluArgIleIleGluGluLeu 532
 QY 521 TTATTGCTCAAAAGCTTTAAATAATCTGTGAGAAAAGGATATACCAAGCTTTTGGAGCC 580
 Db 533 SerAsnIleArgThrValIleGlnIleAlaIleAspGluIleIleArgGlyAlaPheSerSer 552
 QY 581 ATTTTATTTCTAATCTGATCAGTATTTGGCGTCAAGCACTTCTTTCTTCTTCTGATACAA 640

```

Db 553 IlePheAspAlaAsnIleThrThrLeuIleLysValIleIleLeuThrAlaValGlyThr 572
Qy 641 GGGCTATTAAAGGGTTGCTTTGACATGATTTTGAAGATTTTCTTCAATGTTTACG 700
Db 573 GlyAlaIleLysGlyPheAlaIleThrThrGlyIleGlyValAlaThrSerMetPheThr 592
Qy 701 GCTCTTTTCATGACTAAA 718
Db 593 AlaIleValGlyThrArg 598

RESULT 5
US-09-134-001C-5005
; Sequence 5005, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 5005
; LENGTH: 1051
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-5005

Alignment Scores:
Pred. No.: 1,24e-06 Length: 1051
Score: 136.50 Matches: 147
Percent Similarity: 34.12% Conservative: 127
Best Local Similarity: 18.31% Mismatches: 266
Query Match: 4.12% Indels: 263
DB: 4 Gaps: 32

US-09-868-987-1 (1-1864) x US-09-134-001C-5005 (1-1051)
Qy 56 TCTGCAACCGTGGATGGCTGTAGTGTAGTATGCGTTATATGTCAGCAGCCCT 115
Db 273 SerThrAsnSerThrHisMetAlaGln-----LysGlyGlnMetProSer----- 288
Qy 116 ATTTTAAACGTCCTCAAAATCATCCAGTGTCTCAGGGAATTTTACCACCGTGAA 175
Db 289 -----ValProLeuLysAspLeuAlaAspIleSerValGlyAspGluArgThrSer 305
Qy 176 GTGAGCAA----- 184
Db 306 IleSerLysThrAsnGlyLysAspAlaValAsnLeuGlnIleMetLysSerGlnAspAla 325
Qy 185 -----CTCGCTCAGATTAAATCTGGAGCGATGCTCTTTTGTCCGAGGTT 232
Db 326 AsnThrValGlnValAlaArgGluValGlnLysLysValAspGluPheValArgAsnGlu 345
Qy 233 CTCAGTGAAGACGACATCTCTCT-----CATCTTGGGAAA----- 268
Db 346 SerGlyMetLysSerIleLysThrMetAspThrAlaLysProIleGluAspSerLeuTyr 365
Qy 269 -----AAACATGTACACAGGCATTATCTCAGCATGCTGGCTGGCATGCTT 319
Db 366 ThrMetValGluLysAlaAlaLeuGlyThrIle-----ValAlaIleIle 380
Qy 320 ATTGTTTTCATGAGCGTATATATAGATTTTGGAGCGTCATCGCTTCGGAGCTGTTCT 379
Db 381 ValIleLeuLeuPheLeuArgAsnIleArgThrThrAlaIleSerIleValSerIlePro 400
Qy 380 CTGAATCTTTGCTTATCTGGCAGCTCTACAGTATTTGGATGGCCCACTACCTTGTC 439

```

```

Db 401 MetSerIleLeuIleAlaLeuIleAlaLeuLysLeuSerAsnValSerLeuAsnIleLeu 420
Qy 440 CGACTCCTCGGATGTTCTTGTCTATCGGGATGCCGTAGATGCAAAATGTTCTTGTATTC 499
Db 421 ThrLeuGlyAlaLeuThrValAlaIleGlyArgValIleAspAspSerIleValVal 440
Qy 500 GAAAGAATC-----CGAGAGGAATTTTATTTGCTCTCAAGTCTTAAA 541
Db 441 GluAsnIlePheArgArgLeuSerAspProAsnGluLysLeuLysGlyGluAsnLeuIle 460
Qy 542 AAATCTGTAGAAAAAGGATATACCAAGCTTTTGGAGCCATTTTGTATCTTAACCTTGACT 601
Db 461 IleSerAlaThrArg-----GluValPheLysProIleMetSerSerThrLeuVal 477
Qy 602 ACAGTATTGGCCTCAGCACCTCTTTTCTCTAGATACAGGCGCTATT----- 649
Db 478 ThrIleValValPheLeuProLeuValPheVal---SerGlySerValGlyGluMetPhe 496
Qy 650 AAAGGCTTGTCTTGCATTCATTTAGGAATTTTCTTCAATGTTTACGCTCTTTTC 709
Db 497 ArgProPheAlaLeuAlaIleThrPheSerLeuLeuAlaSerLeuValSerIleThr 516
Qy 710 ATG-----ACTAAATTTTCTTCATGCTGTGATGAATAAGACCCCAACAT 754
Db 517 LeuValProSerLeuGlyAlaThrPhePheLysAsnGlyValLysAsnArgGluGlnLys 536
Qy 755 ACACAGTTGCATATGATG----- 772
Db 537 GluGlyLeuGlyThrValGlyArgAlaThrArgSerAlaLeuAsnTrpSerLeuAsnHis 556
Qy 772 ----- 772
Db 557 LysTrpIleValLeuIleValSerIlePheIleLeuValGlySerValIleGlyAla 576
Qy 773 -----AATAAGTTCTGCGGGATAAAGCAT 796
Db 577 ArgAsnLeuGlyThrSerTyrIleSerThrGlyAspAsnLysPheLeuAlaLeuThrTyr 596
Qy 797 -----GAT 799
Db 597 ThrProLysProGlyGluThrGlnLysSerValThrGlnHisAlaGluLysValGlnAsn 616
Qy 800 TTCTTGAGAGATGCAAAAACCTTTGGCTGTT----- 832
Db 617 TyrLeuAspLysLysAspLysValGluThrValGlnTyrSerIleGlyGlyProThrPro 636
Qy 833 -----TCTGAAGTGTCTTTCTTTAGTTCGCTGCT----- 865
Db 637 GlnAspProThrGlySerThrAsnSerMetAlaIleMetIleLysTyrGlnSerAspThr 656
Qy 865 ----- 865
Db 657 ProAsnPheAspGluGluProAspLysValLeuLysHisIleGluThrPheLysGlnPro 676
Qy 866 -----CTCGGTTTGGAGCCTGGAAATTCGTTTTCGGGAATGAT 904
Db 677 GlyGluTrpLysAsnGlnAspLeuGlyThrGlyAlaGlyAsnAsnSerValGluValThr 696
Qy 905 TTTAAAGCA-----GGGTATGCCCTTTTACCTTTTATCCAAAAGACATGCGCATC 952
Db 697 ValLysGlyProAsnThrSerAlaMetLysAspThrValAsnArgValGluLysMetMet 716
Qy 953 AGCGATGTT----- 961
Db 717 ThrAspIleLysGlyIleThrAsnValLysSerAspLeuSerGlnThrTyrAspGlnTyr 736
Qy 962 -----GCTCAAAATGCGT 973
Db 737 GluIleLysValAspGlnAsnLysAlaAlaAspAsnGlyIleSerAlaAlaGlnLeuAla 756
Qy 974 GGCAAGTTGTGCATAAATCTACAGAACTGCTCTTCT----- 1012
Db 757 MetAsnLeuAsnGluAsnLeuProGluLysThrIleSerThrValAsnGluLysGlyLys 776

```

```

QY 1013 TCTAGAGACTTCGGTATT-----CAAACTTTGGATTTCAGAAAGATCAA 1060
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 777 SerIleAspValIleValIleValIleValIleValIleValIleValIleVal 796
QY 1061 ATCTATTATTGATTAAGCTTTAAGCTAATCAAGAGATGAGAGAGAGAGAGAG 1114
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 797 AsnIleLeuAsnIleValIleValIleValIleValIleValIleValIleVal 816
QY 1115 -----CTAAATTAACGATCAATGAGCTGGCTTATTTGGATTGTT 1156
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 817 LysIleSerIleThrProSerIleLeuThrGlnIleAspIleValIleValIleVal 836
QY 1157 GTC---AGAAACAGCGCTAATTTCTCAGCGAAACTCTAAGCAAAAGCAAAATTTGG 1213
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 837 ThrGlyIleValIleThrAspIleValIleValIleValIleValIleValIleVal 853
QY 1214 TCMAAGGTAAAGCAACATTCGAAAGAAATGCGTTATCAGCGAGCAATCGG----- 1267
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 854 AlaIleValIleAsp---LeuGlyIleProSerIleIleValIleValIleValIleVal 872
QY 1268 -----CTTTAGAGCTTTTGCA 1285
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 873 ThrAspIleAspIleValIleValIleValIleValIleValIleValIleValIleVal 892
QY 1286 ATCATCTTGTCTATGATGATTTGCGCTTGAATGCAATATGCT---TTCAGTGGCGTA 1342
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 893 IleValIleLeuIleValIleValIleValIleValIleValIleValIleValIleVal 912
QY 1343 TGGCCTTATTCATGACCTTTGGCTACCGTGACAGCTGTTTATAGCAATTTCTTT 1402
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 913 PheSerIleProPheThrValIleGlyValIleValIleValIleValIleValIleVal 930
QY 1403 TTGAAGAAATTCATAGATTTGCAAGCCATGTGCTTATGATGATGATGATGATGATGAT 1462
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 931 ---GluThrIleSerValPro---SerIleIleGlyMetLeuMet---LeuIleGlyIle 947
QY 1463 TCATTAACAATCTTGTATGATCTTTGATGATGATGATGATGATGATGATGATGATGAT 1522
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 948 ValValIleThrAsnIleValIleValIleValIleValIleValIleValIleValIleVal 967
QY 1523 TTTACCCCTATGATGATTTTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1582
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 968 -----ProMetIleGlnIleValIleValIleValIleValIleValIleValIleVal 985
QY 1583 ATGACACAGCTACATCTATCACTGTTTGTATGATGATGATGATGATGATGATGATGATGAT 1642
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 986 MetThrIleIleIleIleIleIleIleIleIleIleIleIleIleIleIleIleIleIle 1005
QY 1643 GTCCTT-----AATTTTCATTTATGATGATGATGATGATGATGATGATGATGATGAT 1693
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1006 IleLeuIleSerIleGlyMetIleIleIleIleIleIleIleIleIleIleIleIleIle 1025
QY 1694 TCGCTCTTATATGACCACTCTG-----TTGTTGTTTATGATGATGATGATGATGATGAT 1747
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1026 LeuThrIlePheValIleProValIleIleIleIleIleIleIleIleIleIleIleIle 1045
QY 1748 CGCTCAAA 1756
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1046 ThrIleIle 1048

```

RESULT 6
US-09-022-699-2

; Sequence 2, Application US/09022699
; Patent No. 6060277
; GENERAL INFORMATION:
; APPLICANT: KIKLY, KRISTINE
; APPLICANT: SOUTHAN, CHRISTOPHER
; APPLICANT: KNAB, ANNE
; TITLE OF INVENTION: Human AFc1
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: RATNER & PRESTIA

```

; STREET: P. O. BOX 980
; CITY: VALLEY FORGE
; STATE: PA
; COUNTRY: USA
; ZIP: 19482
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: IBM Compatible
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/022,699
; FILING DATE: 12-FEB-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 97304440.7
; FILING DATE: 12-JUN-97
; ATTORNEY/AGENT INFORMATION:
; NAME: PRESTIA, PAUL F
; REGISTRATION NUMBER: 23,031
; REFERENCE/DOCKET NUMBER: GH-70380
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-407-0700
; TELEFAX: 610-407-0701
; TELEX: 846169
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 474 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-022-699-2

```

Alignment Scores:

```

Pred. No.: 0.00143 Length: 474
Score: 117.00 Matches: 55
Percent Similarity: 41.55% Conservative: 31
Best Local Similarity: 26.57% Mismatches: 67
Query Match: 3.53% Indels: 54
DB: 3 Gaps: 9

```

US-09-868-987-1 (1-1864) x US-09-022-699-2 (1-474)

```

QY 1178 CTCTAGCGAAATCTTAACGAAAGCAAAATTTGGTCAAGGTAAACGCAAACTATCG 1237
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 69 LeuIleGlnIleAspIleSerIleThrPheSerPheIleIleValIleValIleValIleVal 83
QY 1238 AAGAAATGCGTTATGACGCGACATCGGCTTTTGAAGCTTTGGCAATCATCTTGTCTG 1297
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 84 -----TyrSerGlnIleGluGlyThrLeu-----IleLeuLeu 94
QY 1298 TATGTAGTTTGGCTTGAATGCA-----TATGCT----- 1330
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 95 PheGlyIleIleProIleIleIleIleIleIleIleIleIleIleIleIleIleIleIle 114
QY 1331 -----TTCAGTGGCGGTATGACGCTTTTATTCATGACCTTTTGGCTTGTGACATC 1381
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 115 GlyProGluIleIleIleIleIleIleIleIleIleIleIleIleIleIleIleIleIle 131
QY 1382 TTGTTTATGACATTTCTTTTGAAGAAATTCAAATATGATTGCAAGCATTTGTGCT 1441
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 132 LeuPheSerIle----- 135
QY 1442 TTAATGCTATTTGGGTATTCATTAACAATATCTTTGATGATGATGATGATGATGATGAT 1501
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 136 ---LeuThrIleLeuProIleIleIleIleIleIleIleIleIleIleIleIleIleIle 154
QY 1502 GAAGATGCCAAGCAACCTGTTTACCCCTATGATGATGATGATGATGATGATGATGATGAT 1561
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 155 PheAsnIleGlnIleIle-----LeuGlyPhePheMetIleIleIleIleIleIle 159
QY 1562 AAGAGCTTACGCGGACGATGATGACAAACAGCTACCACTATACAGTTTGTAAATGCTT 1621

```


QY 1355 CATGACCTTTTGGCTACTGTGCAGCTTGTGTTTATATAGCACATTTCTTTTTCAGAAAAATT 1411
 Db 282 GlyCysLysAlaLysThrCysAspIleLeuTyrlLysAlaGluTyrlPheLeu----- 298
 QY 1415 CAAATAGATTGCAAGCCATTGGTGTCTTTAAATGACTGTATTGGGGTATTTCATTAAACAAT 1474
 Db 299 -----VailLeuAlaValLeuAenSerglyThrAsnPro 309
 QY 1475 ACTTTGATCATTTTTCATCGTATTTCGTGAAGATCGCGAAGCAACCTGTTTACCCCTATG 1534
 Db 310 Ileile----- 311
 QY 1535 CATGTTTCTAGTTAATGATGCCCTTCAAAGACGTTTCAGCCGACGGTA 1582
 Db 312 TyrThrLeuThrAsnLysGluMetArgAlaPheIleArgIleile 327
 RESULT 9
 US-08-760-936-4
 ; Sequence 4, Application US/08760936
 ; Patent No. 5856443
 ; GENERAL INFORMATION:
 ; APPLICANT: MacLennan, A. John
 ; TITLE OF INVENTION: Molecular Cloning and Expression of
 ; TITLE OF INVENTION: G-Protein Coupled Receptors
 ; NUMBER OF SEQUENCES: 14
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Saliwanchik, Lloyd & Saliwanchik
 ; STREET: 2421 N.W. 41st Street, Suite A-1
 ; CITY: Gainesville
 ; STATE: FL
 ; COUNTRY: US
 ; ZIP: 32606
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC Compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/760,936
 ; FILING DATE: December 6, 1996
 ; CLASSIFICATION: 536
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Pace, Doran R.
 ; REGISTRATION NUMBER: 38,261
 ; REFERENCE/DOCKET NUMBER: MAC-100C1
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 352-375-8100
 ; TELEFAX: 352-372-5800
 ; INFORMATION FOR SEQ ID NO: 4:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 383 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 US-08-760-936-4
 Alignment Scores:
 Pred. No.: 0.00459 Length: 383
 Score: 103.50 Matches: 86
 Percent Similarity: 31.80% Conservative: 59
 Best Local Similarity: 18.86% Mismatches: 132
 Query Match: 3.12% Indels: 179
 DB: 2 Gaps: 22
 US-09-868-987-1 (1-1864) x US-08-760-936-4 (1-383)
 QY 248 ATCTCTTCATCTTGGGAAAAACAATGTACAAGCATTTATCTCAGCATGCTGTGGC 307
 Db 40 ValGluLysAspHisGlyIleLysLeuThrSerValValPheIleLeuIleCysCys--- 58
 QY 308 TTGGCAATGCTTATTTGTTTGTGATGAGCGCTATATTATATAGATTGAGGCGCATCGTCTCG 367

QY 1415 CAATGATTTGCAAGCATGTCCTTATGACTGATTTGGGATTTCAATTAACAT 1474
 DB 299 -----ValleuAlaValleuAsnSerGlyThrAspPro 309
 QY 1475 ACTTGATCATTTTGTATGCTATTGCGAAGATCGCAAGCAACCTGTTACCCCTATG 1534
 DB 310 IleIle----- 311
 QY 1535 CATGTTTACTATGATGATGCTTCAAAAGACGTTGACCGCAGCGTA 1582
 DB 312 TyrThrleuThrAsnLysGluMetArgArgAlaPheIleArgIleIle 327
 RESULT 10
 US-08-845-566-3
 ; Sequence 3, Application US/08845566
 ; Patent No. 5912144
 ; GENERAL INFORMATION:
 ; APPLICANT: Au-Young, Janice
 ; APPLICANT: Guegler, Karl
 ; TITLE OF INVENTION: EDG-1 LIKE RECEPTOR
 ; NUMBER OF SEQUENCES: 3
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Incyte Pharmaceuticals, Inc.
 ; STREET: 3174 Porter Drive
 ; CITY: Palo Alto
 ; STATE: CA
 ; COUNTRY: USA
 ; ZIP: 94304
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette
 ; OPERATING SYSTEM: DOS
 ; SOFTWARE: FastSeq for Windows Version 2.0
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/845,566
 ; FILING DATE: Filed Herewith
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER:
 ; FILING DATE:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Billings, Lucy J.
 ; REGISTRATION NUMBER: 36,749
 ; REFERENCE/DOCKET NUMBER: PP-0271 US
 ; TELECOMMUNICATION INFORMATION:
 ; TELEFAX: 415-845-4166
 ; TELEPHONE: 415-855-0555
 ; INFORMATION FOR SRO ID NO: 3:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 381 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; IMMEDIATE SOURCE:
 ; LIBRARY: Genbank
 ; CLONE: 181948
 ; US-08-845-566-3
 Alignment Scores:
 Pred. No.: 0.00888 Length: 381
 Score: 101.00 Matches: 83
 Percent Similarity: 31.83% Conservative: 58
 Best Local Similarity: 18.74% Mismatches: 116
 Query Match: 3.05% Indels: 186
 DB: 2 Gaps: 22
 US-09-868-987-1 (1-1864) x US-08-845-566-3 (1-381)
 QY 290 ATCTCAGATGCTGTGCTTGCAATGCTTATGTTGATGAGCGATATTAATGATTT 349
 DB 53 IleleuIleCysCys-----PheIleIleleuGluAsnIlePheVal----- 66
 QY 350 GGAGGCGTCATCGCTTGGGAGCTGTCTTCTGCAATCTTTGCTATATCTGGGCACT--- 406

DB 67 -----LeuLeuThrIleTrpLysThrLys 74
 QY 407 -----CTACAGATTTTGATGATGCGCACCTACCTTGTGAGCACTGCGTGG 451
 DB 75 LysPheHisArgProMetLysTrpPheIleGlyAsnLeuAlaLeuSerAspLeuAla 94
 QY 452 ATTTGTTCTTGATGGGATGAGCCGTAGATGCAAAATGTTCTGTATGCAAAATCCGA 511
 DB 95 -----GlyValAlaIleThrAlaAsnLeuLeu----- 104
 QY 512 GAGGAATTTTATTTGTCGCAAGTCTTAAAAATCTGTAGAAAAAGATATACCAAGCT 571
 DB 104 ----- 104
 QY 572 TTGGAGCCATTTTGATTTCTAATCTGATACAGATTTGGCCTCAGCACTTCTTTCTTC 631
 DB 105 -----SerGlyAlaThrThrLysLeuThrProAlaGlnTrpPhe 118
 QY 632 CTAGATACAGGCGCTATTAAGGTTTGCTTGACATGATGATTTAGCAATTTTCTTCA 691
 DB 119 LeuArgGluGly-----SerMetPheValAlaLeuSerAlaSer 131
 QY 692 ATGTTTACG-----GCTCTTTGATGACTAAATTTTCTCATGCTGTGATG----- 739
 DB 132 ValPheSerIleuLeuAlaIleAlaIleGluArgTrpIleThrMetLeuLysMetLysLeu 151
 QY 740 --AATTAAGCCACATACACAGTTGCATATGATGATTAAGTTCGGGAGTAAGCAT 796
 DB 152 HisAsnGlySerAsnAsnPheArgLeuPheLeu----- 163
 QY 797 GATTTCTTGAGAGATGCAAAAAAATCTTGAGCTGTTCTGAAAGTGTTCCTTATAGT 856
 DB 164 -----IleSerAlaCys-----TrpValIleSer-----LeuIleGluGly 175
 QY 857 TGGGTGCTCTCGGTTTGAGAGCTGGAATCCGTTTGGGAATGATTTAAAGAGG 916
 DB 176 -----GlyLeuProIleMetGlyTrpAsnCysIle----- 185
 QY 917 TATGCCTTTACTTAAATCCAAAGACATGAGCATGATGTTGCTCAATAGCGTGC 976
 DB 185 ----- 185
 QY 977 AAGTTGTCATTAACATACAGAGCTGTCCTTCTTCTGAGACTTCGATTAACAACA 1036
 DB 186 -----SerAlaLeuSerSer----- 190
 QY 1037 TTGATCTTCAGAAAAATCAAAATCTATTTGATGATTAAGCTTAAAGCTTAAGTAAAG 1096
 DB 191 -----CysSerThrValLeuProLeuTyrHisIleTyrIleLeuPheCysThrThr 208
 QY 1097 CAGATACAGACCTCTGCTCAAAATTAAGCATATGAGCTGGGTTATCTGGATT--- 1153
 DB 209 ValPheThrIleuLeuLeuSerIleValIleLeu-----TyrCysArgIleTyr 225
 QY 1154 ---GTTGTCAGAAACAGGCTAGATTTCTTACAGAAACTCTAAGCAAAAGCAAAATTT 1210
 DB 226 SerIleuValArgThrArgSerArgArgLeu-----ThrPheArgLysAsn----- 240
 QY 1211 TGGTCAAAAGGTAAAGCAAACTATGCAAGAAAAATGCTTATACGCGACCATCGGCTT 1270
 DB 241 -----IleSerLysAlaSerArgSerArgLeuValAlaLeu 253
 QY 1271 TTAGAGCTTTGGCAATCATCTTG---CTCATGTAGTGTGGCTTGAATGAGCAATAT 1327
 DB 254 LeuLysThrValIleIleValLeuSerValPheIleAlaCys-----TrpAlaPro 270
 QY 1328 GCTTTCAGTGGCGTATCGCTTAAATCATGACCTTTTGAGTACCTGTGAGCTTGT 1387
 DB 271 LeuPheIleLeuLeuLeuLeuAspValGlyCysLysValLysThrCysAspIleLeuPhe 290
 QY 1388 ATACACATTTCTTTTGAAGAAAAATTCAAATAGATTTGCAAGCATTTGTCCTTAATG 1447

APPLICANT: CAO, LIANG
APPLICANT: NI, JIAN
APPLICANT: GENTZ, REINER
APPLICANT: BULT, CAROL J.
APPLICANT: SUTTON III, GRANGER G.
APPLICANT: ROSEN, CRAIG A.
TITLE OF INVENTION: Polynucleotides Encoding Human G-Protein
TITLE OF INVENTION: Coupled Receptor GPR1
NUMBER OF SEQUENCES: 30
CORRESPONDENCE ADDRESS:
ADDRESSEE: STERNER, KESSLER, GOLDSTEIN & FOX P.L.L.C.
STREET: 1100 NEW YORK AVE., NW, SUITE 600
CITY: WASHINGTON
STATE: DC
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PATENTIN RELEASE #1.0, VERSION #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/467,947A
FILING DATE: 06-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/04079
FILING DATE: 30-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: STEFFE, ERIC K.
REGISTRATION NUMBER: 36,688
REFERENCE/DOCKET NUMBER: 1488.1140002/EKS/KLM
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-371-2600
TELEFAX: 202-371-2540
INFORMATION FOR SEQ ID NO: 28:
SEQUENCE CHARACTERISTICS:
LENGTH: 381 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: not relevant
MOLECULE TYPE: peptide
US-08-467-947A-28

Alignment Scores:
Pred. No.: 0.00888 Length: 381
Score: 101.00 Matches: 83
Percent Similarity: 31.83% Conservative: 58
Best Local Similarity: 18.74% Mismatches: 116
Query Match: 3.03% Indels: 186
DB: 3 Gaps: 22

US-09-868-987-1 (1-1864) x US-08-467-947A-28 (1-381)

Qy 290 ATCTCAGCATGCTGGCTTGGCAAGCTATTTGTTTATGAGCGGTATATATATATTT 349
Db 53 ileLeuileCysCys-----PheileileLeuGluAsnilePheVal----- 66

Qy 350 GGAGGCGTCATCGCTCGGGAGCTGTTCTTCTGATATCTTTGCTTATCTGGCAGCT--- 406
Db 67 -----LeuLeuthrileTrpIlysthrLys 74

Qy 407 -----CTACAGTATTGGATGCGCCACTCACCTTGTGAGGCTCGCTGGG 451
Db 75 LysPheHisArgProMetTyrTyrPheileGlyAsnLeuAlaLeuSerAspLeuAla 94

Qy 452 ATTGTTCTTGCTATGGGATGCCGTAGATGCAATGTTCTTGTATTCCAAAGATCCGA 511
Db 95 -----GlyValAlaTyrThrAlaAsnLeuLeu----- 104

Qy 512 GAGGAATTTTATTGCTCTCAAGTCTTAAATAATCTGTAGAAAAAGGATATACCAAGCT 571
Db 104 ----- 104

Qy 572 TTGGAGCCATTTTGTGATTTCTAAGTCTACAGTATTTGGCTCAGCAGCTTCTTTTCTTC 631
Db 105 -----SerGlyAlaThrThrTyrLysLeuThrProAlaGlnTrpPhe 118

Qy 632 CTAGATACAGGCGCTATTAAGGCTTTGCTTTGACATTTGATTTAGGAATTTTCTCTTCA 691
Db 119 LeuargGluGly-----SerMetPheValAlaLeuSerAlaSer 131

Qy 692 ATGTTTACG-----GCTCTTTTCATGACTAAATTTTCTTCATGCTGTGGATG----- 739
Db 132 ValPheSerLeuLeuAlaileAlaileGluargTyrileThrMetLeuLysMetLysLeu 151

Qy 740 ---AATAAGCCCAACATACACAGTTGTCATATGATGAATAAGTTTCGTGGGATAAAGCAT 796
Db 152 HisAsnGlySerAsnAsnPheArgLeuPheLeuLeu----- 163

Qy 797 GATTTCTTGAGGATGCAAAAAAATTTTGGCTGTTTCTGGAAGTGTTTTCTTTTAGGT 856
Db 164 -----IleSerAlaCys-----TrpValileSer-----LeuileLeuGly 175

Qy 857 TGCCTTGTCTCTCGGGTTTGGAGCTCGGAATTCCTGTTTGGGAATGGATTTTAAAGGAGG 916
Db 176 -----GlyLeuProIleMetGlyTrpAsnCysile----- 185

Qy 917 TATGCTTTTACCTTTTAAATCCAAAAAGACATGGCATCAGCGATGTTGCTCAAAATCGTGGC 976
Db 185 ----- 185

Qy 977 AAGTTGTGCATAAATACAGGAAGCTGCTTCTTCTTAGAGAGCTTCCGTATTCAAACA 1036
Db 186 -----SerAlaLeuSerSer----- 190

Qy 1037 TTTGATCTTTCAGAAAAGATCAAAATCTATTTTAGTATAAAGCTTTAAAGCTACTATAAG 1096
Db 191 -----CysSerThrValLeuProleuTyrHisLysHisTyrileLeuPheCysThrThr 208

Qy 1097 CAGATACGAGCTCTCTCTTAAATTAACGATCATGAGCTGGCGTTTATTTGGGAT--- 1153
Db 209 ValPheThrLeuLeuLeuLeuSerileValileLeu-----TyrCysArgileTyr 225

Qy 1154 ---GTTGTCAGAAACAGGCTAGATTTCTTACGGAACCTCTAAACGNAACGCAAAATTT 1210
Db 226 SerLeuValArgThrArgSerArgLeu-----ThrPheArgLysAsn----- 240

Qy 1211 TGGTCAAAGGTAACGACCAACTATCGAAGAAATCGTTTATCAGGCGACCATCGGCTT 1270
Db 241 -----IleSerLysAlaSerArgSerSerGluAsnValAlaLeu 253

Qy 1271 TTAGGAGCTTTGGCAATCATCTTG---CTCTATGTGAGTTTTCGCTTTTGAATGGCAAT 1327
Db 254 LeuLysThrValileileValLeuSerValPheileAlaCys-----TrpAlaPro 270

Qy 1328 GCTTTAGTCCGCTATGCGCTTTAATTCATGACCTTTTGGCTACTGTGAGCTTGTGTTT 1387
Db 271 LeuPheileLeuLeuLeuAspValGlyCysLysValLysThrCysAspIleLeuPhe 290

Qy 1388 ATACACATTTCTTTTGAAGAAAATTCMAATAGATTTGCAAGCATTTGGTCTTAAATG 1447
Db 291 ArgAlaGluTyrPheLeu-----ValLeu 298

Qy 1448 ACTGTTTGGGCTTATTCATTAACAATCTTTGATCATTTTGTGATTCGTATTCGTAAGAT 1507
Db 299 AlaValLeuAsnSerGlyThrAsnProIlele----- 309

Qy 1508 CGCCAAGCAACCTGTTTACCCTATGATGATGTTTGTATTAATGATGCCCTTCAAAAGACG 1567
Db 310 -----TyrThrLeuThrAsnLysGluMetArgAla 320

Qy 1568 TTCAGCCGC 1576
Db 321 Pheilearg 323

```

RESULT 14
PCT-US96-10618-4
Sequence 4, Application PC/TUS9610618
GENERAL INFORMATION:
APPLICANT: Coleman, Roger
APPLICANT: Guégier, Karl J.
APPLICANT: Au-Young, Janice
APPLICANT: Bandman, Olga
APPLICANT: Seilhamer, Jeffrey J.
TITLE OF INVENTION: A NOVEL HUMAN EDG-2 RECEPTOR HOMOLOG
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US96/10618
FILING DATE: 20-JUN-1996
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/000,352
FILING DATE: 20-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/567,817
FILING DATE: 06-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: Glaister, Debra J.
REGISTRATION NUMBER: 33,888
REFERENCE/DOCKET NUMBER: PR-0042 PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 381 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
IMMEDIATE SOURCE:
LIBRARY: GenBank
CLONE: 119130
PCT-US96-10618-4
Alignment Scores:
Pred. No.: 0.00888 Length: 381
Score: 101.00 Matches: 83
Percent Similarity: 31.83% Conservative: 58
Best Local Similarity: 18.74% Mismatches: 116
Query Match: 3.05% Indels: 186
DB: Gaps: 22
US-09-868-987-1 (1-1864) x PCT-US96-10618-4 (1-381)
Oy 290 ATCTAGAGATGCTGGCTTGCATGCTTATTTGATGACGATATATATATATTT 349
Db ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
53 lleuilellecysCys-----PhelelleleugluuallieleVal----- 66
Oy 350 GGAGCGTCATCGCTTCGAGACTGTTCTTGATCTTTTGCTTATCTGGGAGCT--- 406
Db 67 -----leuenuhrllelletpLysThrLys 74
Oy 407 -----CTACGATATTGGATGCGGCACCTGACCTTGTGACGACTGCGTGGG 451
::: ||| ||||| ||||| |||||

```

[illegible]


```

Db 415 -----AlaserValserThrValLeuSerLeuLeu 425
Qy 1616 ATGCTTTGTTATAGCGCCTCTGTC-----TTAATTTGCAATTATT 1663
Db 426 PheValIleIleuGlnValAlaValLeuLysTyrTyrArgPheAsnArgIleSerLeu 445
Qy 1664 ATGACCAATAGGATCTCTTGAACACTTATCGTCTTTATATGACCACTCTGTG 1723
Db 446 PheIleValLysLeuIleLeuGlyMetIleMetSerIleValAlaGlnThrValMet 465
Qy 1724 TTGTTATGCGCGTAA 1741
Db 466 LeuAlaLeuProSerLys 471

RESULT 16
US-09-542-733-2
; Sequence 2, Application US/09542733
; Patent No. 632333
; GENERAL INFORMATION:
; APPLICANT: Ping Tsui
; TITLE OF INVENTION: Mouse EDG1
; FILE REFERENCE: GP-70610
; CURRENT APPLICATION NUMBER: US/09/542,733
; EARLIER FILING DATE: 2000-04-04
; EARLIER APPLICATION NUMBER: 60/127,696
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 382
; TYPE: PRT
; ORGANISM: MUS MUSCULUS
US-09-542-733-2

Alignment Scores:
Pred. No.: 0.0225 Length: 382
Score: 97.50 Matches: 86
Percent Similarity: 31.57% Conservative: 57
Best Local Similarity: 18.98% Mismatches: 131
Query Match: 2.94% Indels: 179
Gaps: 22
DB: 4

US-09-868-987-1 (1-1864) x US-09-542-733-2 (1-382)
Qy 257 GATCTTGGAATAAACAATGTACACAGGCAATATCTCAGACGCTGTGGCTGGCAATG 316
Db 42 AsphIsGlyIleLysLeuThrSerValPheIleLeuIleCysCys----- 57
Qy 317 CTATTGTTGATGAGCGGTATATTATGATTTGGAGCGTCATCGCTTGGAGACTGT 376
Db 58 PheIleIleLeuGlnAsnIlePheVal----- 66
Qy 377 CTTCGATCTTTTGCTTATCGGCAGCT-----CTACAGATTTG 418
Db 67 -----LeuLeuThrIleTyrLysThrLysPheHisArgProMetTyrThr 83
Qy 419 GATCGCCACTCAGCTGTCTAGAGCTCGTGGATTTGTTGCTANGAGGAGCCGTA 478
Db 84 IleGlyAsnLeuAlaLeuSerAspLeuAla-----GlyValAlaLys 98
Qy 479 GATGCAATGTCTTGTATTTGAAAGAAATCCGAGAGGAATTTTATGTCTCAAGTCT 538
Db 99 ThrIaAsnLeuLeu----- 104
Qy 539 AAAAAATGTGAGAAAAAGATATACCAAGGCTTTTGAGGCAATTTGATCTCACTTG 598
Db 105 -----SerGlyAla 107
Qy 599 ACTACAGTATTGGCTCAGCACTTCTTTCTCTCAGATACAGGCGCTATTAAAGGTTT 658
Db 108 ThrThrTyrLysLeuThrPheAlaGlnThrPheLeuAlaGlyLys----- 122

```

```

Qy 659 GCTTGACATGATTTAGGAATTTCTCTCAATGTTACG-----GCTCTTTCATG 712
Db 123 -----SerMetPheValAlaLeuSerAlaSerValPheSerLeuAlaIleAlaIle 140
Qy 713 ACTAAATTTTCTTCATGCTGTGATGAATTAAGACCAACATACAGTTGATGATG 772
Db 141 GluArgTyrIleThrMetLeuLysMetLys-----LeuHisAsnGlySerAsnSer 158
Qy 773 AATTAAGTTCGTGGGATTAAGCATGATTTCTTGAGAGATGCAAAAACCTTGGCGT 832
Db 159 ArgSerPheLeu-----LeuIleSerAlaCys-----TrpValIle 170
Qy 833 TCTGAAGTGTCTTTCTTTAGGTGCGTTGCTCGCGGTTTGAGACCTGATTCGTT 892
Db 171 Ser-----LeuIleLeuGly-----GlyLeuProIleMetGlyTrpAsnCysIle 185
Qy 893 TTGGGAATGATTTTAAAGAGGATGCTTACCTTTAATCCAAAGACATGCGATC 952
Db 185 ----- 185
Qy 953 AGCGATGTCCTCAATGCGTGCAAAAGTTTGCAATTAACAGAGAGCTGCTTCT 1012
Db 186 -----SerSerLeuSer 189
Qy 1013 TCTAGAGACTTCGTAATTCAAACATTTGATCTTCAGAAAAGATCAAAATCTATTAGT 1072
Db 190 Ser-----CysSerThrValLeuProLeuTyrHisLys 200
Qy 1073 GATTAAGCTTTAAGCTATATACAGATACGAGCTCTCTCTTAATTAACAGATCATG 1132
Db 201 HisTyrIleLeuPheCysThrThrValPheThrLeuLeuLeuSerIleValIleLeu 220
Qy 1133 AGCTGGCTTATGTGGAT-----GTTTCAGAAACAGCGCTGATTTCTCAACGA 1186
Db 221 -----TyrCysArgIleTyrSerLeuValAlaTyrThrSerArgLysLeu----- 235
Qy 1187 AACTTAAGCAACCAAAATTTGTCMAAGGTACAGCAACCAACTATGAAAGAAATG 1246
Db 236 ThrPheArgLysAsnIleSerLysAlaSerArgSerSerLys----- 250
Qy 1247 CGTTATACGCGCAACATCGGCTTTTGAAGACTTTGGCAATCATCTTG---CTCTATG 1303
Db 251 -----SerLeuAlaLeuLeuLysThrValIleIleValLeuSerValPheIle 266
Qy 1304 AGTTGGCTTTGAATRGCAATATGCTTTGAGTCCGCTATGCGCTTAATTCATGACT 1363
Db 267 AlaCys-----TrpAlaProLeuPheIleLeuLeuLeuAspValGlyCysLys 283
Qy 1364 TTGGCTACCTGTGCAGCTGTGTTATAGACATTTCTTTTGAAGAAATTCAAATAGAT 1423
Db 284 AlaLysThrCysAspIleLeuTyrLysAlaGluTyrPheLeu----- 297
Qy 1424 TTGCAGCCATGTGCTTAAATGATGATTTAGCTTATGAGGATTCATTAACATCTGATC 1483
Db 298 -----ValLeuAlaValLeuAsnSerGlyThrAsnProIleIle--- 310
Qy 1484 ATTTTGAATGATGTGTGTAAGATCGCAAGCAACCTGTTACCTCATGACATGTTTA 1543
Db 311 -----TyrThrLeu 313
Qy 1544 GTTAATGATCCCTTCAAAAGAGTTTCAGCGCAGCGTA 1582
Db 314 ThrAsnLysGluMetArgAlaPheIleArgIleVal 326

RESULT 17
US-07-998-289B-8
; Sequence 8, Application US/07998289B
; Patent No. 6027876
; GENERAL INFORMATION:
; APPLICANT: Black, Bruce C
; APPLICANT: Taylor, Martin
; APPLICANT: Heckel, David G
; TITLE OF INVENTION: Method for Monitoring Pesticide

```



```

, TITLE OF INVENTION: Resistance
, NUMBER OF SEQUENCES: 40
, CORRESPONDENCE ADDRESS:
, ADDRESSEE: Darby & Darby PC
, STREET: 805 Third Avenue
, CITY: New York
, STATE: New York
, COUNTRY: US
, ZIP: 10022
, COMPUTER READABLE FORM:
, MEDIUM TYPE: Floppy disk
, COMPUTER: IBM PC compatible
, OPERATING SYSTEM: PC-DOS/MS-DOS
, SOFTWARE: Patent In Release #1.0, Version #1.30
, CURRENT APPLICATION DATA:
, APPLICATION NUMBER: US/07/998,289B
, FILING DATE: 30-DEC-1992
, CLASSIFICATION: 435
, ATTORNEY/AGENT INFORMATION:
, NAME: Robinson, Joseph R
, REGISTRATION NUMBER: 33,448
, REFERENCE/DOCKET NUMBER: 0646/OA939
, TELECOMMUNICATION INFORMATION:
, TELEPHONE: 212-527-7700
, TELEFAX: 212-753-6237
, TELEX: 236687
, INFORMATION FOR SEQ ID NO: 8:
, SEQUENCE CHARACTERISTICS:
, LENGTH: 1820 amino acids
, TYPE: amino acid
, STRANDEDNESS: single
, TOPOLOGY: linear
, MOLECULE TYPE: protein
, US-07-998-289B-8

```

Alignment Scores:		
Pred. No.:	0.0514	Length:
Score:	97.50	Matches:
Percent Similarity:	33.58%	Conservative:
Best Local Similarity:	17.92%	Mismatches:
Query Match:	2.94%	Indels:
DB:	3	Gaps:
		24
		1820

US-09-868-987-1 (1-1864) x US-07-998-289B-8 (1-1820)

Qy	287	ATTATCTCAGACGCTGTGGCTTGGCAATGCCTTATTGTTTGATGAGCGTATATATATAGA	346
Db	1319	LeuthrAsnAlaIrcpCysTrpLeuAspPheValIleValMetValSerLeuIleAsnPhe	1338
Qy	347	TTTGGAGGCGTCATCGCTTCGGAGCGTGTCTTCGAATCTTTGCTTATCTCGGCGACCT	406
Db	1339	valAserLeuValGlyAlaGlyIle-----GlnAla	1350
Qy	407	CTACAGTATTGGATCGGCCACTCACCTTGTTCAGACCTCGCTGGGATTTGTTCTTGCTATG	466
Db	1351	PheLysThrMetArgThrLeuAlaAlaLeuArgProLeuArgAlaMetSerArgMetCln	1370
Qy	467	GGATGCGCGTAGATCCAAATGTTCTTGATTTCGAAAGAAATCCGAGAGAAATTTTATTG	526
Db	1371	GlyMetArgValValValAsnAlaLeuVal-----	1380
Qy	527	TCTCAAAGTCTTAATAAATCTGTAGAAAAAGGATATACCAAGGCTTTTTGGAGCATTTTT	586
Db	1381	-----GlnAlaIleProserIlePhe	1387
Qy	587	GATTCTAACTTGACTACAGTATTGGCCTCAGACATCTTTTCTTCTCATAGATACAGGCCT	646
Db	1388	Asn-----ValLeuLeuValCysLeuIlePheTrpLeu-----	1398
Qy	647	ATTAAAGGGTTGCTTTTGACATTGATTTAGAAATTTCTCTTCAATGTTTACGGCTCTT	706
Db	1399	-----IlePheAlaIleMetGlyValGlnLeu-----	1407

DB 1693 LeuLeuPheAlaLeuAlaMetSerLeuProAlaLeuPheAsnIleCys----- 1708

QY 1706 ATGCACACACCTCTGTGTGTGTATGATC 1735
 |||||
 |||||
 DB 1709 -----LeuLeuLeuPheLeuVal 1714

RESULT 18
 US-08-808-793-23
 ; Sequence 23, Application US/08808793
 ; Patent No. 5858713
 ; GENERAL INFORMATION:
 ; APPLICANT: Soderlund, David M.
 ; APPLICANT: Ingles, Patricia J.
 ; TITLE OF INVENTION: CALCIUM PERMEABLE INSECT SODIUM CHANNELS
 ; TITLE OF INVENTION: ANT USE THEREOF
 ; NUMBER OF SEQUENCES: 32
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP
 ; STREET: Clinton Square, P.O. Box 1051
 ; CITY: Rochester
 ; STATE: New York
 ; COUNTRY: USA
 ; ZIP: 14603
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/808,793
 ; FILING DATE:
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 60/034,361
 ; FILING DATE: 24-DEC-1996
 ; CLASSIFICATION: 435
 ; INFORMATION NUMBER: 435
 ; NAME: Bitman, Susan J.
 ; REGISTRATION NUMBER: 34,103
 ; REFERENCE/DOCKET NUMBER: 19603/1062 (D-1906A)
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 716-263-1636
 ; TELEFAX: 716-263-1600
 ; INFORMATION FOR SEQ ID NO: 23:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 2100 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: not relevant
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-808-793-23

Alignment Scores:
 Pred. No.: 0.0554 Length: 2100
 Score: 97.50 Matches: 95
 Percent Similarity: 33.58% Conservativeness: 83
 Best Local Similarity: 17.92% Mismatches: 171
 Query Match: 2.94% Indels: 181
 DB: 2 Gaps: 24

US-09-868-987-1 (1-1864) x US-08-808-793-23 (1-2100)

QY 287 ATATCTCAGCATGCTGTG3CTTGSCAATGCTTATGTTTGATGAGCGTATATATATAGA 346
 ::::|
 ::::|
 DB 1332 LeuThrAsnAlaTrpCysTrpLeuAspPheValIleValMetValSerIleuIleAsnPhe 1351
 ::::|
 ::::|

QY 347 TTGGAGCGTCATCGCTTGGAGACCTTCTCTCGAATCTTTGCTTATCTGGGACACT 406
 ::::|
 ::::|

DB 1352 ValAlaSerLeuValGlyAlaGlyIle-----GlnAla 1363

QY 407 CTACAGTATTTGGATGCGCCACTACCTTGTTCAGACCTCGTGAGATTCCTTGTCTATG 466
 ::::|
 ::::|
 DB 1364 PheLysThrMetArgThrLeuArgAlaLeuArgProLeuArgAlaMetSerArgMetGln 1383
 ::::|
 ::::|
 QY 467 GGGATGCGCGTATGATGCAATGTTCTTGATTCGAAAGATCCGAGAGATTTTATATG 526
 |||||
 |||||
 DB 1384 GlyMetArgValValValAsnAlaLeuVal----- 1393

QY 527 TCTCAAGTCTTAAATAATCTGTAGAAAAAGATATACCAAGCTTTGGAGCCATTTT 586
 ::::|
 ::::|
 DB 1394 -----GlnAlaIleProSerIlePhe 1400

QY 587 GATCTTAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 646
 ::::|
 ::::|
 DB 1401 Asn-----ValLeuLeuValCysLeuIlePheTrpLeu----- 1411

QY 647 ATTAAGGTTTGCTTTGACATTTAGGAATTTCTTCATGTTTACGGCTCTT 706
 ::::|
 ::::|
 DB 1412 -----IlePheAlaIleMetGlyValGlnLeu 1420

QY 707 TTCATGACTAAATTTTCTTCATGCTGTGATGATGATGATGATGATGATGATGATGATGAT 766
 |||||
 |||||
 DB 1421 PheAlaGlyLysTrpPheLysCysGluAspMetAsnGlyThrLys----- 1435

QY 767 ATGATGAATTAAGTCTGGGATTAAGCATGATTTCTTGAGAGATGCAAAAACCTTTGG 826
 ::::|
 ::::|
 DB 1436 -----LeuSerHisGluIleIleProAsnArgAsn----- 1445

QY 827 GCTGTTTCTGAAGTGTCTTTCTTTAGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 886
 |||||
 |||||
 DB 1446 -----AlaCysGluSerGluAsnTrpThrTrpValAsn 1456

QY 887 TCCGTTTGGGAATGATTTTAAAGA---GGGATGCTTTACCTTAAATCCAAAGAG 943
 |||||
 |||||
 DB 1457 SerAlaMetAsnPheAspHisValGlyAsnAlaTrpLeuCysLeuPhe----- 1472

QY 944 CATGCGATCAGCATGTTGCTCAAAATGCGTGAAGTTGTGATTAACCTACAGAAAGCT 1003
 |||||
 |||||
 DB 1473 -----GlnValAlaThrPheLysGlyTrpIle-----GlnIleMetAsnAsp 1486

QY 1004 GGTCTTCTCTGACAGACTCCGATTTCAAAACATTTGATCTTCAGAAAATCAAAATC 1063
 ::::|
 ::::|
 DB 1487 AlaIleAspSerArgGluValAlaLeuLysGlnProIle-----ArgGluThrAsnIle 1503

QY 1064 TATTTAGTATGAACCTTTAAGCTTACTACAGATACAGACCTCTCTCTTAAATTA 1123
 |||||
 |||||
 DB 1504 TyrMetTrpLeuTrpPheValPhePheIle-----IlePheGlySerPhePheThrLeu 1521

QY 1124 ACGATCAGAGCTGGCGTATTTGGATTTGTGCAAGAACGGGCTAGATTTCTCTAC 1183
 ::::|
 ::::|
 DB 1522 AsnLeu-----PheIleGlyValIleIleAspAsn-----Phe 1532

QY 1184 GGAACCTTAACGAAACGCA----- 1204

DB 1533 AsnGluGlnLysLysLysAlaGlyLysSerLeuGluMetCysPheMetThrGluAspGlnLys 1552

QY 1205 AAATTTTGTCAAGATGACGACCAAACTATCGAAGAA----- 1243

DB 1553 LysTrpTrpSerAlaMetLysLysMetGlySerLysLysProLeuLysAlaIleProArg 1572

QY 1244 -----ATGCGTATACGGGACCATCGGCTTTAGAGACT-----TTGGCA 1285

DB 1573 ProArgTrpArgProGlnAlaIleValPheGluIleValThrAspLysLysPheAspIle 1592

QY 1286 ATCATCTGCTCATGAGATTTG-----CGCTTGAATG 1321
 |||||
 |||||
 DB 1593 IleIleMetLeuPheIleGlyLeuAsnMetPheThrMetThrLeuMetAspArgTrpAspAla 1612

QY 1322 CAATATGCTTTCAGTGGCGATGCGCTTAAATTCATGACCTTTTG----- 1366

DB 1613 SerAspThrTrpAsnAlaValLeuAspTrpLeuAsnAlaIlePheValIlePheSer 1632

QY 1367 GCTACCTGTCAGCTCTTTATAGCA-----CATTCTTTTGAAGAAATTCAA 1417
 Db 1633 SerGluCysLeuLeuLysLeuPheAlaLeuArgTyrHisTyrPheIleGluProTyrPheAsn 1652
 QY 1418 ATAGATTGTCAGCAGCTGGTCTTAATGACTGATGTTGGGTATTCATTAACAATACT 1477
 Db 1653 Leu---PheAspValValValIleLeuSerIleLeuGly----- 1665
 QY 1478 TTGATCATTTTGTGATGCGTATGCGTGAAGATCGCCAAAGCAACCTGTTTACCCCTATGCAT 1537
 Db 1665 ----- 1665
 QY 1538 GTTTAGTTATGATGCCCTTCAAAGAGCTTC---ACCCGACGGTAATGACAAAGCT 1594
 Db 1666 LeuValLeuSerAspIleGluLysTyrPheValSerProTyrLeuLeuArgValVal 1685
 QY 1595 ACAACTCTATCAGTTTCTTAATGCTTTTGTATAGCGGCTCTCT- - - - -GTC 1645
 Db 1686 ArgValAlaLysValGlyArgValLeuArgLeuValLysGlyAlaLysGlyIleArgThr 1705
 QY 1646 TTTAATTTGTCATTATTATGACCATAGGAGTCTCTCTAGGAACCTTTATCGTCTCTTAT 1705
 Db 1706 LeuLeuPheAlaLeuAlaMetSerLeuProAlaLeuPheAsnIleCys- - - - - 1721
 QY 1706 ATTGCACCACCTCTGTTGTTGTTATGTC 1735
 Db 1722 -----LeuLeuLeuPheLeuVal 1727

RESULT 19

US-08-772-512A-19
 ; Sequence 19, Application US/08772512A.
 ; Patent No. 6022705
 ; GENERAL INFORMATION:
 ; APPLICANT: Soderlund, David M.
 ; APPLICANT: Knipple, Douglas C.
 ; APPLICANT: Ingles, Patricia J.
 ; TITLE OF INVENTION: INSECT SODIUM CHANNELS FROM
 ; TITLE OF INVENTION: INSECTICIDE-SUSCEPTIBLE AND INSECTICIDE-RESISTANT HOUSE
 ; TITLE OF INVENTION: FLIES
 ; NUMBER OF SEQUENCES: 19
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: NIXON, HARGRAVE, DEVANS & DOYLE LLP
 ; STREET: P.O. Box 1051, Clinton Square
 ; CITY: Rochester
 ; STATE: New York
 ; COUNTRY: USA
 ; ZIP: 14603
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/772,512A
 ; FILING DATE:
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/608,618
 ; FILING DATE: 01-MAR-1996
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Braman, Susan J.
 ; REGISTRATION NUMBER: 34,103
 ; REFERENCE/DOCKET NUMBER: 19603/601 (CRFD-1657)
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 716-263-1636
 ; TELEFAX: 716-263-1600
 ; INFORMATION FOR SEQ ID NO: 19:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 2100 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: not relevant
 ; TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-772-512A-19

Alignment Scores:
 Pred. No.: 0.0554 Length: 2100
 Score: 97.50 Matches: 95
 Percent Similarity: 33.58% Conservatives: 83
 Best Local Similarity: 17.92% Mismatches: 171
 Query Match: 2.94% Indels: 181
 DB: 3 Gaps: 24

US-09-868-987-1 (1-1864) x US-08-772-512A-19 (1-2100)

QY 287 ATTATCTCAGCATGCTGCTGGCAATGCTTATTTTGTAGCGGTATATATAGA 346
 Db 1332 LeuThrAsnAlaTyrCysTyrPheValSerLeuIleAsnPhe 1351
 QY 347 TTGGAGCGCTCATCGCTTCGGGAGCTGTTCTTCTGAATCTTTTGTCTATCTGGGAGCT 406
 Db 1352 ValAlaSerLeuValGlyAlaGlyGlyIle-----GlnAla 1363
 QY 407 CTACAGTATTGTCATGCGCCACTCACCTTGTTCAGGACTCGCTGGGATTTCTTGTCTATG 466
 Db 1364 PheLysThrMetArgThrLeuArgAlaLeuArgProLeuArgAlaMetSerArgMetGln 1383
 QY 467 GGGATGCGCTAGATGCAAAATGTTCTTGTATTCGAAAGAATCCGAGAGGAATTTTATTG 526
 Db 1384 GlyMetArgValValValAsnAlaLeuVal----- 1393
 QY 527 TCTCAAAGTCTTAAATAAATCTGTAGAAAAGGATATACCAAGGCTTTTGGAGCCATTTT 586
 Db 1394 -----GlnAlaIleProSerIlePhe 1400
 QY 587 GATTCTAACTTGACTACAGATGTTGGCTCAGCACTCTTTTCTCTCAGATACAGGGCCT 646
 Db 1401 Asn-----ValLeuLeuValCysLeuIlePheTyrPheLeu----- 1411
 QY 647 ATTAAAGGTTGCTTTGACATGATTTTAGGAATTTCTCTCAATGTTTACGGCTCTT 706
 Db 1412 -----IlePheAlaIleMetGlyValGlnLeu 1420
 QY 707 TTCATGACTAAATTTTCTTCATGCTGTGATGATGAATGAAGACCCCAACATACACATTCGAT 766
 Db 1421 PheAlaGlyLysTyrPheLysCysGluAspMetAsnGlyThrLys----- 1435
 QY 767 ATGATGAATAAGTTCGTGGGATAAAGCATGATTTCTGTAGAGGATGCAAAAACCTTTGG 826
 Db 1436 -----LeuSerHisGluIleIleProAsnArgAsn----- 1445
 QY 827 GCTGTTTCTGGAAGTGTGTTTCTTTTAGTTGCGTTCCTCTCGGGTTTGGAGCCTGGAAT 886
 Db 1446 -----AlaCysGluSerGluAsnTyrThrTyrPheValAsn 1456
 QY 887 TCCGTTTGGGAATGGATTTTAAAGGA---GGGTATGCCCTTTACCTTTTAAATCCAAAAGAG 943
 Db 1457 SerAlaMetAsnPheAspHisValGlyAsnAlaTyrLeuCysLeuPhe----- 1472
 QY 944 CATGGCATCAGCGATGTTGCTCAAAATCGGTGCAAAAGTGTGCAATAAACTACAGGAAGCT 1003
 Db 1473 -----GlnValAlaThrPheLysGlyTyrPheIle-----GlnIleMetAsnAsp 1486
 QY 1004 GGCTTTTCTAGAGACTTCCCGTATTTCAAAACATTTTGGATCTTTCAGAAAAGATCAAAATC 1063
 Db 1487 AlaIleAspSerArgGluValAspLysGlnProIle-----ArgGluThrAsnIle 1503
 QY 1064 TATTTTAGTGAATAAGCTTTAAGCTATACCTAAGCAGATACAGAGCTCTCTCTCTCTCTCTCT 1123
 Db 1504 TyrMetTyrLeuTyrPheValPhePheIle-----IlePheGlySerPhePheThrLeu 1521
 QY 1124 ACGATCATGAGCTGGCGTATTGCGGATTTGTTCTCAGAAACAGCGCTAGATTTCTCTAC 1183
 Db 1522 AsnLeu-----PheIleGlyValIleIleAspAsn-----Phe 1532


```

APPLICATION NUMBER: US 60/005,498
FILING DATE: 04-OCT-1995
ATTORNEY/AGENT INFORMATION:
NAME: Highlander, Steven L.
REGISTRATION NUMBER: 37,642
REFERENCE/DOCKET NUMBER: UTSD:450
TELECOMMUNICATION INFORMATION:
TELEPHONE: (512) 418-3000
TELEFAX: (512) 474-7577
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1090 amino acids
STRANDEDNESS:
TYPE: amino acid
TOPOLOGY: linear
US-08-726-214-4

Alignment Scores:
Pred. No.: 0.0447 Length: 1090
Score: 97.00 Matches: 54
Percent Similarity: 37.98% Conservative: 44
Best Local Similarity: 20.93% Mismatches: 102
Query Match: 2.93% Indels: 59
Gaps: 11

US-09-868-987-1 (1-1864) x US-08-726-214-4 (1-1090)
QY 237 GTGAGAGAGCATCTCTTGTGATCTTGGAAAAAC-----AATGTACACAGGCAATTA 290
Db 575 Ileglnaaglieserleuleuphetyrasmlysaasnilegulysglutyargalathr 594
QY 291 TCT-----CAGCATGCTGGCTTGGCAATGCTTATGTTGTGATGAGCGTAT 338
Db 595 AlaleuprolaaphelystyrytvalthrCysAlaCysleullepheleucysilephe 614
QY 339 ATTATAGATTG-----GAGGCGCA 359
Db 615 IilevalglnilleuvalleuprolysThrserileuuglyPheSerPheGlyAlaAla 634
QY 360 TCGCTTCGGAGAGCTGTTCTTGTGATCTTGTTCGTTATCTGGGAGCTCTACAGTATTTGG 419
Db 635 PheleuserleullepheleullephevalCysphealaglylnleuGln----- 652
QY 420 ATGGGCCACTGCTGTCAGAGCTCGCTGGAGATGTTCTTGGTATG----- 467
Db 653 -----CysSerlylsalaserThrserleumetIpleuLeuLysSer 667
QY 468 -----GAGTGGCGCA-----GATGCAAT 487
Db 668 SerglylleileAlaAsnaryproItyrproargileSerleuthrIlevalThrThra 687
QY 488 GTTCTTGATTGAAAGATCCGAGAGGAAATTTTATTGTCTCAAGCTTAAAAATCT 547
Db 688 IilelleuthrMetAlaValPheAsnmetPhepheleuserAsnserGlnGluThrThr 707
QY 548 GTAGAAAAAGATATCCACAGGCTTTGGAGCCATTTTGTATTCATCTACATACAGTA 607
Db 708 LeuproThraIAsnthrsersAsnAlaAsnValSerValProAspAsnGlnAlaSerile 727
QY 608 TTGGCCTCAGACCTCTTCTTCTCTCCTAGATACAGGCGCTATTAAGGTTTGGCTTGA 667
Db 728 LeuhisAlaArgasnLeuphePheleuproItyr-----PheIleTyser 742
QY 668 TTGATTTAGGAATTTCTCT--TCAATGTTACGGCTCTT--TTCATGACTAAATTT 721
Db 743 CyslleuuglyleuilleserCysSerValPheleuAargvalAsnItyrGlnleuLysmet 762
QY 722 TTCTTCATGCTG-----TGATGATAAGACCCCAACATACACAGTTCAT 766
Db 763 LeuilemetMetValAlaValGlytyrAsnThrIleLeuLeuHisThrHisAlaHis 782
QY 767 ATGATGAATAAGTTGCTGG3GATA-----AAGCATGATTTCTTGAAGAGATGCAAA 817

```

```

Db 783 ValLeuAspAlaItyrSerGlnValleuPheGlnArgProGlyIleItyrLysAspleuLys 802
QY 818 AAACCTTGCGCTCTTTCGAGAGTGTCTTTAGTTGCGTCTCTCGGG 871
Db 803 ThmeGlySerValserleuserIlePhePheIleThrleuLeuValleuGly 820

RESULT 23
PCT-US95-11808-3
Sequence 3, Application PC/TUS9511808
GENERAL INFORMATION:
APPLICANT: Iyengar, Srinivas Ravi V.
TITLE OF INVENTION: MUTANT ACTIVATED GALPHA AND ADENYLYL
TITLE OF INVENTION: CYCLASE 2 FOR USE AS THERAPEUTIC AGENTS
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Brumbaugh, Graves, Donohue and
STREET: 30 Rockefeller Plaza
CITY: New York
STATE: New York
COUNTRY: U.S.
ZIP: 10112-0228
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/11808
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/307,896
FILING DATE: 16-SEP-1994
ATTORNEY/AGENT INFORMATION:
NAME: Clark, Richard S.
REGISTRATION NUMBER: 26,154
REFERENCE/DOCKET NUMBER: 29970 165/28755
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 408-2500
TELEFAX: (212) 765-2519
TELEX: 650 6111063
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1090 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US95-11808-3

Alignment Scores:
Pred. No.: 0.0447 Length: 1090
Score: 97.00 Matches: 54
Percent Similarity: 37.98% Conservative: 44
Best Local Similarity: 20.93% Mismatches: 102
Query Match: 2.93% Indels: 59
Gaps: 11

US-09-868-987-1 (1-1864) x PCT-US95-11808-3 (1-1090)
QY 237 GTGAGAGAGCATCTCTTGTGATCTTGGAAAAAC-----AATGTACACAGGCAATTA 290
Db 575 Ileglnaaglieserleuleuphetyrasmlysaasnilegulysglutyargalathr 594
QY 291 TCT-----CAGCATGCTGGCTTGGCAATGCTTATGTTGTGATGAGCGTAT 338
Db 595 AlaleuprolaaphelystyrytvalthrCysAlaCysleullepheleucysilephe 614
QY 339 ATTATAGATTG-----GAGGCGCA 359
Db 615 IilevalglnilleuvalleuprolysThrserileuuglyPheSerPheGlyAlaAla 634

```

```
Qy 360 TCGCTTCGGGAGCTGTTCTTCTGAATCTTTTGGCTTATCTGGGCACTCTACAGTATTGG 419
Db 635 PheLeuSerLeuIlePheIleLeuPheValCysPheAlaGlyGlnLeuLeuGln----- 652
Qy 420 ATGCGCCACTCACCTGTGTGTCAGGACTCGCTGGGATTGTTCTTCTGCTATGG----- 467
Db 653 -----CysSerIlysIysAlaSerThrSerLeuMetTrpLeuLeuIysSer 667
Qy 468 -----GGATGGCCGTA-----GATGCAAAAT 487
Db 668 SerGlyIleIleAlaAsnArgProTrpProArgIleSerLeuThrIleValThrAla 687
Qy 488 GTTCTTGATTGCAAGAAGATCCGAGAGAAATTTTATGCTCTCAAGCTTAAAAAATCT 547
Db 688 IleIleLeuThrMetAlaValPheAsnMetPhePheLeuSerAsnSerGluGluThrThr 707
Qy 548 GTAGAAAAGGATATACCAAGGCTTTTGGAGCCATTTTGTATCTTAACTTGACTACAGTA 607
Db 708 LeuProThrAlaAsnThrSerAsnAlaAsnValSerValProAspAsnGlnAlaSerIle 727
Qy 608 TTGGCCTCAGCATTCTTTCTTCTCTAGATACAGGCGCTATTAAAGGTTTGTCTTTGACA 667
Db 728 LeuHisAlaArgAsnLeuPhePheLeuProTyr-----PheIleTyrSer 742
Qy 668 TTGATTTAGGAATTTCTCT---TCAATGTTTACGGCTCTT---TTCACTGACTAAATTT 721
Db 743 CysIleLeuGlyLeuIleSerCysSerValPheLeuArgValAsnTyrGluLeuIysMet 762
Qy 722 TTCCTTCATGCTG-----TGGATGAATAAGACCAACATACATACAGTTGCTAT 766
Db 763 LeuIleMetMetValAlaLeuValGlyTyrAsnThrIleLeuLeuHisThrHisAlaHis 782
Qy 767 ATGATGAATAAGTTCGTGGGATA-----AAGCATGATTTCTTGAGAGATGCAAA 817
Db 783 ValLeuAspAlaTyrSerGlnValLeuPheGlnArgProGlyIleTrpLysAspLeuLys 802
Qy 818 AACTTTGGGCTGTTTCTGGAGTCTTTTCTTTAGTGGTGGCTGCTCGG 871
Db 803 ThrMetGlySerValSerLeuSerIlePhePheIleThrLeuLeuValLeuGly 820

RESULT 24
US-09-262-477-2
; Sequence 2, Application US/09262477A
; Patent No. 6423508
; GENERAL INFORMATION:
; APPLICANT: George Livi
; APPLICANT: Derk Bergema
; APPLICANT: Jeffrey Stadel
; APPLICANT: Winnie Chan
; APPLICANT: Randall Johnson
; APPLICANT: Shelagh Wilson
; APPLICANT: Jon Chambers
; APPLICANT: Phillippe Robert
; APPLICANT: Nassirah Khandoudi
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDE SEQUENCES OF HUMAN EDG-1C AND
; TITLE OF INVENTION: METHODS OF SCREENING FOR AGONISTS AND ANTAGONISTS FOR HUMAN
; TITLE OF INVENTION: EDG-1C RECEPTOR AND SPHINGOSINE 1-PHOSPHATE LIGAND AND
; TITLE OF INVENTION: INTERACTION THEREOF
; FILE REFERENCE: GP50006
; CURRENT APPLICATION NUMBER: US/09/262,477A
; CURRENT FILING DATE: 1999-03-04
; EARLIER APPLICATION NUMBER: 60/077,369
; EARLIER FILING DATE: 1998-03-09
; EARLIER APPLICATION NUMBER: 60/087,102
; EARLIER FILING DATE: 1998-05-28
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 382
; TYPE: PRT
; ORGANISM: Human
US-09-262-477-2
```

```
Alignment Scores:
Pred. No.: 0.0293 Length: 382
Score: 96.50 Matches: 83
Percent Similarity: 32.05% Conservative: 59
Best Local Similarity: 18.74% Mismatches: 116
Query Match: 2.91% Indels: 185
DB: 4 Gaps: 22

US-09-868-987-1 (1-1864) x US-09-262-477-2 (1-382)
Qy 290 ATCTCAGCATGCTGCTGGCTGGCAATCTTATGTTTGTAGAGCGTATATTATAGATTT 349
Db 53 IleuIleCysCys-----PheIleIleuGluAsnIlePheVal----- 66
Qy 350 GGAGGCGTCATCGCTTCGGGAGCTGTTCTTCTGAATCTTTTGTCTTATCTGGGAGCT--- 406
Db 67 -----LeuLeuThrIleTrpLysThrLys 74
Qy 407 -----CTACAGTATTGTGGATCGGCCACTCACCTTGTACAGACTCGCTGGG 451
Db 75 LysPheHisArgProMetTyrTyrPheIleGlyAsnLeuAlaLeuSerAspLeuLeuAla 94
Qy 452 ATTGTTCTGTCTATGGGATGGCCGTAGATGCAAAATGTTCTTGTATTTCGAAAGAAATCGA 511
Db 95 -----GlyValAlaTyrThrAlaAsnLeuLeu----- 104
Qy 512 GAGGAATTTTATTGTTCTCAAAAGTCTTAAAAAATCTGTAGAAAAAGGATATACCAAGGCT 571
Db 104 ----- 104
Qy 572 TTGGAGCCATTTTGTGATTTCTAATTCGACTACAGTATTTGGCTCAGCAGCTTCTTTTCTTC 631
Db 105 -----SerGlyAlaThrThrTyrLysLeuThrProAlaGlnTrpPhe 118
Qy 632 CTAGATACAGGCGCTATTAAAGGTTTTCCTTGGACATTGATTTTAGGAATTTTCTCTTCA 691
Db 119 LeuArgGluGly-----SerMetPheValAlaLeuSerAlaSer 131
Qy 692 ATGTTTACG-----GCTCTTTTCATGACTAAATTTTTCCTCATGCTGTGGATG----- 739
Db 132 ValPheSerLeuLeuAlaIleAlaIleGluArgTyrIleThrMetLeuLysMetLysLeu 151
Qy 740 ---AATAAGACCCCAACATACACAGTTGCATATGATGAATTAAGTTCGTGGGATAAAGCAT 796
Db 152 HisAsnGlySerAsnAsnPheArgLeuPheLeuLeu----- 163
Qy 797 GATTCTTGAGAGATGCAAAAACATTTTGGCTGTTTCTTCTGGAAGTGTTTTCTTTTAGGT 856
Db 164 -----IleSerAlaCys-----TrpValIleSer-----LeuIleLeuGly 175
Qy 857 TGCCTGCTCTCGGGTTTGGAGCCTCGAAATTCGCTTTTGGGATGGATTTTAAAGGAGGG 916
Db 176 -----GlyLeuProIleMetGlyTrpAsnCysIle----- 185
Qy 917 TATGCTTTTACCTTTTAATCCAAAAGACATGGCATCAGCGATGTTGTCAAATCGTGGC 976
Db 185 ----- 185
Qy 977 AAAGTTGTGCATAAACTACAGGAAGCTGGTCTTCTTCTAGAGACTTCCGTTATTCAAAACA 1036
Db 186 -----SerAlaLeuSerSer----- 190
Qy 1037 TTTGGATCTTCAGAAAAGATCAAAATCTATTATTTAGTATAAGCTTTTAAGCTATATAAG 1096
Db 191 -----CysSerThrValLeuProLeuTyrHisLysHisTyrIleLeuPheCysThrThr 208
Qy 1097 CAGATACAGGCTCTCTCTTAAATAATTAAACGATCATGAGCTGGCGTTATTGTGGGAT--- 1153
Db 209 ValPheThrLeuLeuLeuLeuSerIleValIleLeu-----TyrCysArgIleTyr 225
Qy 1154 ---GTTGTACAGAACACAGGCTAGATTTTCTACGGAACCTCTAAACGAAACCGCAAAATTT 1210
Db 226 SerLeuValArgThrArgSerArgArgLeu-----ThrPheArgLysAsnIleSerLys 243
```


QY 1193 AACGAAACGCAAAATTTTGGTCAAGGTAGCAGCAAACTATCGAAGAAATCGCTTAT 1252
 Db 944 -----PheTrpGluGlnTyrMetThrLeuArgSerSerLeuAlaMetile 958
 QY 1253 CAGCGACCATCGGCTTTTAGGAGCTTTGGCAATCATC---TTGCTCTATGAGTTTG 1309
 Db 959 LeuAlaCysVal---LeuLeuAlaLeuValLeuValSerLeuLeuLeuSerVal 977
 QY 1310 CGCTTTGAATGCAATATGCTTTCAGTCCGTATGCGCTTTAAATTCATGACCTTTTGCT 1369
 Db 978 -----Trp-----AlaAlaValLeuValLeuSerValLeuAla 989
 QY 1370 ACCTGTGAGTCTGCTTTTATAGCAATTTCTTTTGAAGAAATTCATAGATTGCAA 1429
 Db 990 SerLeuAlaGlnIlePheGlyAla-----MetThrLeuLeuGlyIleLeuSer 1006
 QY 1430 GCCATTGCTGCTTTAAATGACTGTATGCGGTATTCATTAACAATACATCTTTCATTTT 1489
 Db 1007 AlaIleProAlaValIleLeuIleLeuSerValGlyMet-----MetLeuCysPhe 1023
 QY 1490 GATCGTATTCGTAGATCGCAAGCAACCTGTTTACCCTATGCAATGTTTATGTTAAT 1549
 Db 1024 -----AsnValLeuIleSer 1028
 QY 1550 GATGCCCTTCAAAGACGTTTCAGC-----CGCAGCGTAATGACACAGCTACA 1597
 Db 1029 LeuGlyPheMetThrSerValGlyAsnArgGlnArgValGlnLeuSerMetGlnMet 1048
 QY 1598 ACTGTATCAGTTTGTGTTAATGCTTTTCTATAGCGGCTCCTCTGCTTTT----- 1648
 Db 1049 SerLeuGlyProLeuValHisGlyMetLeuThrSerGlyValAlaValPheMetLeuSer 1068
 QY 1649 -----AATTTGCAATTTATGACCATAGGATTTCTT 1681
 Db 1069 ThrSerProPheGluPheValIleArgHisPheCysTrpLeuLeuValValLeuCys 1088
 QY 1682 CTAGAACTTTATCGTCTCTTTTATATGACCACTCTGTTGTTTATGCTCGGTAAA 1741
 Db 1089 ValGlyAlaCysAsnSerLeuValPhe---ProIleLeuLeuSerMetValGlyPro 1107
 QY 1742 GAAATCGCTCAAAATAGTACCGTTAACTTAATCTAACGTGTAGCAATATAAAATCT 1801
 Db 1108 Glu-----AlaGluLeuValProLeuGluHisProAspArgIleSerThr---ProSer 1124
 QY 1802 CCTTGGACCTTTAGTCCCAAGG 1825
 Db 1125 ProLeuProValArgSerSerLys 1132

RESULT 31

US-08-954-668-6
 ; Sequence 6, Application US/08954668
 ; Patent No. 6172200
 ; GENERAL INFORMATION:
 ; APPLICANT: SCOTT, MATHEW P
 ; APPLICANT: GOODRICH, LISA V
 ; APPLICANT: JOHNSON, RONALD L
 ; TITLE OF INVENTION: Patched Genes and their Use
 ; NUMBER OF SEQUENCES: 19
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Foley, Hoag & Eliot
 ; STREET: One Post Office Square
 ; CITY: Boston
 ; STATE: MA
 ; COUNTRY: USA
 ; ZIP: 02109
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: ASCII(text)
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/954,668
 ; FILING DATE: 20-Oct-1997

; CLASSIFICATION: 530
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Vincent, Matthew P.
 ; REGISTRATION NUMBER: 36709
 ; REFERENCE/DOCKET NUMBER: SUV-003.06
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 617-832-1000
 ; TELEFAX: 617-832-7000
 ; INFORMATION FOR SEQ ID NO: 6:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1285 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-954-668-6
 Alignment Scores:
 Pred. No.: 0.123 Length: 1285
 Score: 93.50 Matches: 62
 Percent Similarity: 42.91% Conservative: 53
 Best Local Similarity: 23.13% Mismatches: 90
 Query Match: 2.82% Indels: 63
 DB: 4 Gaps: 14

US-09-868-987-1 (1-1864) x US-08-954-668-6 (1-1285)

QY 1082 TTAAGCTATATAGCATAGATACGAGCCTCTCTCTAAATAATTAACGATCATGAGCTGGCGT 1141
 Db 908 LeuThrAspThrSerGlnIleLysThrLeuIleGlyHisIleArgAspLeuSerValLys 927
 QY 1142 TATGTGCGGATGTTGTCAGAAACAGCCCTAGA-----TTTCTCTACGGAACCTCT 1192
 Db 928 TyrGluGlyPheGlyLeuProAsnTyrProSerGlyIleProPheIle----- 943
 QY 1193 AACGAAACGCAAAATTTTGGTCAAGGTAGCAGCAAACTATCGAAGAAATCGCTTAT 1252
 Db 944 -----PheTrpGluGlnTyrMetThrLeuArgSerSerLeuAlaMetile 958
 QY 1253 CAGCGACCATCGGCTTTTAGGAGCTTTGGCAATCATC---TTGCTCTATGAGTTTG 1309
 Db 959 LeuAlaCysVal---LeuLeuAlaLeuValLeuValSerLeuLeuLeuSerVal 977
 QY 1310 CGCTTTGAATGCAATATGCTTTCAGTCCGTATGCGCTTTAAATTCATGACCTTTTGCT 1369
 Db 978 -----Trp-----AlaAlaValLeuValLeuSerValLeuAla 989
 QY 1370 ACCTGTGAGTCTGTTTATAGCAATTTCTTTTGAAGAAATTCATAGATTGCAA 1429
 Db 990 SerLeuAlaGlnIlePheGlyAla-----MetThrLeuLeuGlyIleLeuSer 1006
 QY 1430 GCCATTGCTGCTTTAAATGACTGTATGCGGTATTCATTAACAATACATCTTTCATTTT 1489
 Db 1007 AlaIleProAlaValIleLeuIleLeuSerValGlyMet-----MetLeuCysPhe 1023
 QY 1490 GATCGTATTCGTAGATCGCAAGCAACCTGTTTACCCTATGCAATGTTTATGTTAAT 1549
 Db 1024 -----AsnValLeuIleSer 1028
 QY 1550 GATGCCCTTCAAAGACGTTTCAGC-----CGCAGCGTAATGACACAGCTACA 1597
 Db 1029 LeuGlyPheMetThrSerValGlyAsnArgGlnArgValGlnLeuSerMetGlnMet 1048
 QY 1598 ACTGTATCAGTTTGTGTTAATGCTTTTCTATAGCGGCTCCTCTGCTTTT----- 1648
 Db 1049 SerLeuGlyProLeuValHisGlyMetLeuThrSerGlyValAlaValPheMetLeuSer 1068
 QY 1649 -----AATTTGCAATTTATGACCATAGGATTTCTT 1681
 Db 1069 ThrSerProPheGluPheValIleArgHisPheCysTrpLeuLeuValValLeuCys 1088
 QY 1682 CTAGAACTTTATCGTCTCTTTTATATGACCACTCTGTTGTTTATGCTCGGTAAA 1741
 Db 1089 ValGlyAlaCysAsnSerLeuValPhe---ProIleLeuLeuSerMetValGlyPro 1107

Db 1089 ValGlyAlaCysaenSerLeuValPhe---ProIleLeuLeuSerMetValGlyPro 1107
 Qy 1742 GAAATGCTCAAAATAAGTACCGTTAACTTAATCTTAACGTTAGCAATATAAAATCT 1801
 Db 1108 Glu-----AlaGluLeuValProLeuGluHisProAspArgIleSerThr---ProSer 1124
 Qy 1802 CCTTGGGACTTTAGTCCCAAGG 1825
 Db 1125 ProLeuProValArgSerSerLys 1132

RESULT 32
 US-08-918-658-6
 ; Sequence 6, Application US/08918658
 ; Patent No. 6423154
 ; GENERAL INFORMATION:
 ; APPLICANT: SCOTT, MATHEW P
 ; GOODRICH, LISA V
 ; JOHNSON, RONALD L
 ; TITLE OF INVENTION: Patched Genes and their Use
 ; NUMBER OF SEQUENCES: 19
 ; CORRESPONDENCE ADDRESSES:
 ; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
 ; STREET: Four Embarcadero Center, Suite 3400
 ; CITY: San Francisco
 ; STATE: CA
 ; COUNTRY: US
 ; ZIP: 94111
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; OPERATING SYSTEM: IBM PC compatible
 ; SOFTWARE: Patentin Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/918,658
 ; FILING DATE: 22-Aug-1997
 ; CLASSIFICATION: <Unknown>
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/656,055
 ; FILING DATE: <Unknown>
 ; APPLICATION NUMBER: 08/540,406
 ; FILING DATE: <Unknown>
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Rowland, Bertram I
 ; REGISTRATION NUMBER: 20015
 ; REFERENCE/DOCKET NUMBER: a60190-1
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 415-781-1989
 ; TELEFAX: 415-358-3249
 ; INFORMATION FOR SEQ ID NO: 6:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1285 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 6:
 ; US-08-918-658-6

Alignment Scores:
 Pred. No.: 0 123 Length: 1285
 Score: 93.50 Matches: 62
 Percent Similarity: 42.91% Conservative: 53
 Best Local Similarity: 23.13% Mismatches: 90
 Query Match: 2.82% Indels: 63
 DB: 4 Gaps: 14

US-09-868-987-1 (1-1864) x US-08-918-658-6 (1-1285)
 Qy 1082 TTAAGCTACTAAGACAGATGAGCCCTCTCTCTAAATTAACGATCATGAGCTGGCGT 1141
 Db 908 LeuThrAspTrnSerGlnIleLysThrLeuIleGlyHisIleArgAspLeuSerValLys 927
 Qy 1142 TATTGGGATTGTTGTTCAGAAACAGGCTAGA-----TTTCTCTACGAAACTCT 1192

Db 928 TyrGluGlyPheGlyLeuProAsnTrpProSerGlyIleProPheIle----- 943
 Qy 1193 AAACGAAACGCAAAATTTGGTCAAGTAAGCAGCAAACTATCGAAGAAATGCGTTAT 1252
 Db 944 -----PheTrpGluGlnTrpMetThrLeuArgSerSerLeuAlaMetIle 958
 Qy 1253 CAGCGACCATCGGGCTTTTAGAGCTTTTAGCAATATC---TTGGCTATAGTAATTGG 1309
 Db 959 LeuAlaCysVal---LeuLeuAlaAlaLeuValSerLeuLeuLeuSerVal 977
 Qy 1310 CGCTTGAATGCAATATGCTTTCAGTCCGTATGCGCTTATATTCATGACCTTTGGCT 1369
 Db 978 -----Trp-----AlaAlaValLeuValIleLeuSerValLeuAla 989
 Qy 1370 ACCTGTCAGCTCTGTATTATAGCAGATCTTTTGAAGAAATTCAAATAGATTGCA 1429
 Db 990 SerLeuAlaGlnIlePheGlyAla-----MetThrLeuLeuGlyIleLysLeuSer 1006
 Qy 1430 GCCATTGTCCTTAAATGACTGATTTGGGGTATTCATTAACAATACTTTGATCATTTT 1489
 Db 1007 AlaIleProAlaValIleLeuIleLeuSerValGlyMet-----MetLeuCysPhe 1023
 Qy 1490 GATCGTATTGTGAAGATGCGCAAGCAGACTGTATACCCCTATGATGATGTTAGTTAAT 1549
 Db 1024 -----AsnValLeuIleSer 1028
 Qy 1550 GATGCCCTTCAAAAAGCGTTACG-----CGCAGGTAATGACACAGCTTAC 1597
 Db 1029 LeuGlyPheMetThrSerValGlyAsnArgGlnArgArgValGlnLeuSerMetGlnMet 1048
 Qy 1598 ACTCTATACGTTTGTATATGCTTTTGTATAGGGGCTCCCTCTCTCTTT----- 1648
 Db 1049 SerLeuGlyProLeuValHisGlyMetLeuThrSerGlyValAlaValPheMetLeuSer 1068
 Qy 1649 -----AATTTGCAATTAATTAATGCAATGAGGATTTCTT 1681
 Db 1069 ThrSerProPheGluPheValIleArgHisPheCysTrpLeuLeuValValLeuCys 1088
 Qy 1682 CTAGGAACCTTATATGCTCTCTTATTAATGACACACCTGTGTGTTGTTATGTCGTAA 1741
 Db 1089 ValGlyAlaCysaenSerLeuValPhe---ProIleLeuLeuSerMetValGlyPro 1107
 Qy 1742 GAAATGCTCAAAATAAGTACCGTTAACTTAATCTTAACGTTAGCAATATAAAATCT 1801
 Db 1108 Glu-----AlaGluLeuValProLeuGluHisProAspArgIleSerThr---ProSer 1124
 Qy 1802 CCTTGGGACTTTAGTCCCAAGG 1825
 Db 1125 ProLeuProValArgSerSerLys 1132

RESULT 33
 PCT-US95-13233-6
 ; Sequence 6, Application PC/TUS9513233
 ; GENERAL INFORMATION:
 ; APPLICANT: THE BOARD OF TRUSTEES OF TEH LELAND STANFORD JUNIOR UNIVERSITY
 ; TITLE OF INVENTION: Patched Genes and their Use
 ; NUMBER OF SEQUENCES: 19
 ; CORRESPONDENCE ADDRESSES:
 ; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
 ; STREET: Four Embarcadero Center, Suite 3400
 ; CITY: San Francisco
 ; STATE: CA
 ; COUNTRY: US
 ; ZIP: 94111
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; OPERATING SYSTEM: IBM PC compatible
 ; SOFTWARE: Patentin Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: PCT/US95/13233
 ; FILING DATE: 06-OCT-1990

[illegible][illegible]

QY 1550 GATGCCCTTCAAAAAGAGCTTACG-----CGACGGTAATGACACAGCTACA 1597
 Db 1030 LeuGlyPheMetThSerValGlyAsnArgGlnArgValGlnLeuSerMetGlnMet 1049
 QY 1598 ACTGTATCAGATTGTTGTTAATGCTTTTGTATAGCGGCTCCTGTCTTT----- 1648
 Db 1050 SerLeuGlyProLeuValInISGlyMetLeuThSerGlyValAlaValPheMetLeuSer 1069
 QY 1649 -----AATTTGCATTATTAATGACCATAGGATTTCTT 1681
 Db 1070 ThrSerProPheGluPheValIleArgHisPheCysTrpLeuLeuValIleValCys 1089
 QY 1682 CTAGGAACCTTATCGTCTTTATATGCAACCCCTGTGTTGTTATAGTCCGTA 1741
 Db 1090 ValGlyAlaCysAsnSerLeuValPhe--ProIleLeuLeuSerMetValGlyPro 1108
 QY 1742 GAAATATGCTCAAAATAAGACGCTTAACCTTAATCTAAGCTGAGCATATGAAAAATCT 1801
 Db 1109 Gln-----AlaGluLeuValProLeuGlnHisProAsnArgIleSerThr--ProSer 1125
 QY 1802 CCTTTGGAGCTTTAGTCCCAAGG 1825
 Db 1126 ProLeuProValArgSerSerLys 1133

RESULT 35
 US-08-460-900C-62
 ; Sequence 62, Application US/08460900C
 ; Patent No. 6165747

GENERAL INFORMATION:
 APPLICANT: Ingham, Phillip W.
 APPLICANT: McMahon, Andrew P.
 APPLICANT: Tabin, Clifford J.
 APPLICANT: Bumcrot, David A.
 APPLICANT: Marti-Gorosua, Elisa
 TITLE OF INVENTION: Vertebrate Embryonic Pattern-Inducing
 TITLE OF INVENTION: Proteins and Uses Related Thereto
 NUMBER OF SEQUENCES: 62
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: FOLEY, HONG & ELIOT LLP
 STREET: One Post Office Square
 CITY: Boston
 STATE: MA
 COUNTRY: USA
 ZIP: 02109

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/460,900C
 FILING DATE: 5-JUNE-1995
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/435,093
 FILING DATE: 4-MAY-1995
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/356,060
 FILING DATE: 14-DEC-1994
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/176,427
 FILING DATE: 30-DEC-1993

ATTORNEY/AGENT INFORMATION:
 NAME: Vincent, Matthew P.
 REGISTRATION NUMBER: 36,709
 REFERENCE/DOCKET NUMBER: HMV-006.05
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (617) 832-1000
 TELEFAX: (617) 832-7000
 INFORMATION FOR SEQ ID NO: 62:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1299 amino acids
 TYPE: amino acid
 TOPOLOGY: linear

MOLECULE TYPE: protein
 US-08-460-900C-62

Alignment Scores:
 Pred. No.: 0.124
 Score: 93.50
 Percent Similarity: 42.91%
 Best Local Similarity: 23.13%
 Query Match: 2.82%
 DB: 4
 Gaps: 14

US-09-868-987-1 (1-1864) x US-08-460-900C-62 (1-1299)

QY 1082 TTAAGCTATATAGAGAGATAGAGAGCTCTCTCAAAATTAACGATGACGTGGCGT 1141
 Db 922 LeuThAsPthSerGlnIleLysThrLeuIleGlyHisIleArgAspLeuSerValLys 941
 QY 1142 TATTGGGAGATTGTGTGCAAAACAGGCTTAA-----TTTCTTAACGAAACTCT 1192
 Db 942 TyrGluGlyPheGlyLeuProAsnTyrProSerGlyIleProPheIle----- 957
 QY 1193 AAACGAAAGCAAAATTTGGTCAAGGTAGAGAGCAACTATGCAAGAAATGCGTTAT 1252
 Db 958 -----PheTrpGluGlnTyrMetThrLeuArgSerSerLeuAlaMetIle 972
 QY 1253 CAGGACCATCGGCGCTTTTAAAGAGCTTTGGCAATCATC---TTGCTTATGTAGTTTG 1309
 Db 973 LeuAlaCysVal---LeuLeuAlaIleValIleValSerLeuLeuLeuSerVal 991
 QY 1310 CGCTTGAATGCGCATATGCTTCAAGTCCGTAAGCGCTTAAATTCATGACCTTTGGCT 1369
 Db 992 -----Trp-----AlaAlaValIleValIleLeuSerValIleAla 1003
 QY 1370 ACCTGTGACAGTCTTGTATTATGACATTTCTTTGAAGAAATTCAAATAGATTGGCA 1429
 Db 1004 SerLeuAlaGlnIlePheGlyAla-----MetThrLeuLeuGlyIleLysLeuSer 1020
 QY 1430 GCCATTGCTCTTTAAATGACTGATTGGGATTCATTAACAAATCTTGTATCATTTT 1489
 Db 1021 AlIleProAlaValIleLeuIleLeuSerValGlyMet-----MetLeuCysPhe 1037
 QY 1490 GATCGATTGCGAAGATCGCAAGCGAAGCTGTATACCCCTATGCATGTTTAGTTAAT 1549
 Db 1038 -----AsnValIleLeuIleSer 1042
 QY 1550 GATGCCCTTCAAAAAGAGCTTACG-----CGACGGTAATGACACAGCTACA 1597
 Db 1043 LeuGlyPheMetThSerValGlyAsnArgGlnArgValGlnLeuSerMetGlnMet 1062
 QY 1598 ACTGTATCAGATTGTTGTTAATGCTTTTGTATAGCGGCTCCTGTCTTT----- 1648
 Db 1063 SerLeuGlyProLeuValInISGlyMetLeuThSerGlyValAlaValPheMetLeuSer 1082
 QY 1649 -----AATTTGCATTATTAATGACCATAGGATTTCTT 1681
 Db 1083 ThrSerProPheGluPheValIleArgHisPheCysTrpLeuLeuValIleValCys 1102
 QY 1682 CTAGGAACCTTATCGTCTTTATATGCAACCCCTGTGTTGTTATAGTCCGTA 1741
 Db 1103 ValGlyAlaCysAsnSerLeuValPhe--ProIleLeuLeuSerMetValGlyPro 1121
 QY 1742 GAAATATGCTCAAAATAAGACGCTTAACCTTAATCTAAGCTGAGCATATGAAAAATCT 1801
 Db 1122 Gln-----AlaGluLeuValProLeuGlnHisProAsnArgIleSerThr--ProSer 1138
 QY 1802 CCTTTGGAGCTTTAGTCCCAAGG 1825
 Db 1139 ProLeuProValArgSerSerLys 1146

RESULT 36
 US-08-674-509B-48
 ; Sequence 48, Application US/08674509B
 ; Patent No. 6261786


```

; GENERAL INFORMATION:
; APPLICANT: Ingham, Phillip W.
; APPLICANT: McMahon, Andrew P.
; APPLICANT: Tabin, Clifford J.
; APPLICANT: Marigo, Valeria
; TITLE OF INVENTION: SCREENING ASSAYS FOR HEDGEHOG AGONISTS
; TITLE OF INVENTION: AND ANTAGONISTS
; NUMBER OF SEQUENCES: 48
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FOLEY, HOAG & ELIOT LLP
; STREET: One Post Office Square
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109-2170
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/674,509B
; FILING DATE: 02-JUL-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/460,900
; FILING DATE: 05-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Vincent, Matthew P.
; REGISTRATION NUMBER: 36,709
; REFERENCE/DOCKET NUMBER: HMV-006.06
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-832-1000
; TELEFAX: 617-832-7000
; INFORMATION FOR SEQ ID NO: 48:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1299 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-674-509B-48

Alignment Scores:
Pred. No.: 0.124
Score: 93.50
Percent Similarity: 42.91%
Best Local Similarity: 23.13%
Query Match: 2.82%
DB: 4

US-09-868-987-1 (1-1864) x US-08-674-509B-48 (1-1299)
Qy 1082 TTAAGCTATACATAGCAGATACGAGCCTCTCTCTTAAATTAACGATCATGCTGGGT 1141
Db 922 LeuthraspThrSerGlnIleuThrLeuIleuGlyHisIleArgAspLeuSerValuys 941
Qy 1142 TATTGTGGGATTGTGTGAGAAAGCGCTAGA-----TTTCTACGGAACCTCT 1192
Db 942 TyrGluGlyPheGlyLeuProAsnTyrProSerGlyIleProPheIle----- 957
Qy 1193 AACGAAACGCAAAATTTTGTCAAGTAGAGCAGCAACTATCGAAGAAATGCGTTAT 1252
Db 958-----PheTrpGluGlnThrMetThrLeuArgSerSerLeuAlaMetile 972
Qy 1253 CAGCGACCATCGGCTTTTAGGAGCTTTGGCAATCATC---TTGCTCTATGTGATTG 1309
Db 973 LeuAlaCysVal---LeuLeuAlaLeuValLeuValSerLeuLeuLeuSerVal 991
Qy 1310 CGCTTTGAATGGCAATATGCTTTTCAGTCGCGGTATCGCTTTAAATTCATGACCTTTGGCT 1369
Db 992-----Ttp-----AlaAlaValLeuValIleLeuSerValLeuAla 1003
Qy 1370 ACCTGTGCGACTCTGTTTATAGCACATTTCTTTTGAAGAAAAATTCAAATAGATTGCAA 1429

```

```

Db 1004 SerLeuAlaGlnIlePheGlyAla-----MetThrLeuLeuGlyIleLysLeuSer 1020
Qy 1430 GCCATTGGTCTTTAAATGACTGTATTGGGGTATTTCATTAACAATACCTTTGATCATTTT 1489
Db 1021 AlaIleProAlaValIleLeuIleLeuSerValGlyMet-----MetLeuCysPhe 1037
Qy 1490 GATCGTATTCGTGAAGATCGCCAAAGCAACCTGTTTACCCCTATGCATGTTTGTAGTTAA 1549
Db 1038-----AsnValLeuIleSer 1042
Qy 1550 GATGCCCTTCAAAAGACGTTTCAGC-----CGCACGGTAATGACACACAGCTACA 1597
Db 1043 LeuGlyPheMetThrSerValGlyAsnArgGlnArgValGlnLeuSerMetGlnMet 1062
Qy 1598 ACTCTATCAGCTTTTGTAAATGCTTTTGTATAGCGCGCTCTCTGCTTT----- 1648
Db 1063 SerLeuGlyProLeuValHisGlyMetLeuThrSerGlyValAlaValPheMetLeuSer 1082
Qy 1649-----AATTTGCATTTATTATGACCATAGGAGTCTT 1681
Db 1083 ThrSerProPheGluPheValIleArgHisPheCysTrpLeuLeuValValLeuCys 1102
Qy 1682 CTAGGAACCTTTATCGTCTCTTTATATTGCACCACTCTGTTGTTGTTTATGTCGCTGTA 1741
Db 1103 ValGlyAlaCysAsnSerLeuLeuValPhe---ProIleLeuLeuSerMetValGlyPro 1121
Qy 1742 GAAATCGCTCAAAATAAGTACCGTTAAACTTAATCTAAGCTGTAGCAATATAAAATCT 1801
Db 1122 Glu-----AlaGluLeuValProLeuGluHisProAspArgIleSerThr---ProSer 1138
Qy 1802 CTTTGGGACTTTTAGTCCCAAGG 1825
Db 1139 ProLeuProValArgSerSerLys 1146

RESULT 37
US-08-954-698-48
; Sequence 48, Application US/08954698
; Patent No. 6271363
; GENERAL INFORMATION:
; APPLICANT: Ingham, Phillip W.
; APPLICANT: McMahon, Andrew P.
; APPLICANT: Tabin, Clifford J.
; TITLE OF INVENTION: Vertebrate Embryonic Pattern-Inducing
; TITLE OF INVENTION: Proteins and Uses Related Thereto
; NUMBER OF SEQUENCES: 48
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FOLEY, HOAG & ELIOT LLP
; STREET: One Post Office Square
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109-2170
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/954,698
; FILING DATE: 20-OCT-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/462,386
; FILING DATE: 05-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/435,093
; FILING DATE: 04-MAY-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/356,060
; FILING DATE: 14-DEC-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/176,427
; FILING DATE: 30-DEC-1993

```

US-09-134-001C-5567

Sequence 5567, Application US/09134001C
Patent No. 6380370
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
FILE REFERENCE: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
CURRENT APPLICATION NUMBER: US/09/134,001C
PRIORITY FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 5567
LENGTH: 344
TYPE: PRT
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-5567

Alignment Scores:

Pred. No.:	0.0702	Length:	344
Score:	93.00	Matches:	73
Percent Similarity:	37.08%	Conservative:	69
Best Local Similarity:	19.06%	Mismatches:	133
Query Match:	2.81%	Indels:	108
DB:	4	Gaps:	17

US-09-868-987-1 (1-1864) x US-09-134-001C-5567 (1-344)

OY	452 ATTTGCTTGCATGGAGTGCGCGTAGATGCAAAATGTTCTGTATTGGAAGAATCCGA	511
Db	44 leuileuSerMetSerleuSerleuSerTrpaSpValGlyIysIeuAspValPhe	63
OY	512 GAGGATTTTATTTGTCCTCAA--AGTCTTAATAAAAATCTGTAGAAAAAGCATTPACCAAG	568
Db	64 AsnGlnThrIleLeuPheGlnValArgIeuProArgLeuLeuGluIuIaMetIeuThrGly	83
OY	569 GCCTTT-----GGAGCCATTTTGTGATTCCTAAGTACTGACACTACAGTATGGCCTCA	616
Db	84 AlaValIeuThrLeuAlaGlyIenValTyrgInIleValIeuAsnAsnProIeuAlaAsp	103
OY	617 GCACCTTTTCTTCTTCATGATACAGGCGCTTATTAAGGGTTGCTTGGACATTGATTTA	676
Db	104 SerPheThrIeuGlyIeuAlaSerGlyAlaSerIeuGlySerGlyIleAlaIeuPheIeu	123
OY	677 GGA-----ATTTCTCT-----TCAATGTTTACGGCTCTT	706
Db	124 GlyIeuSerPheIeuTriPheProIlePheSerIleIlePheSerIeuIeThrIeuIeu	143
OY	707 TTCATGACTAAATTTTTCTTCATGCTGTGTGATGAATTAAGACCACAATACACAGTTGCAT	766
Db	144 IeuValIeuSerValSerIaMetIeu-----	152
OY	767 ATGATGATTAAGTTCGTGGGATTAAGCATGATTTCTTGAGAGATGCAAAAAAATTGG	826
Db	152 -----	152
OY	827 GCTGTTTTCGAAGTGTTTTCTTTTAGGTTCGTGCTCTCGGGTTTGAGCC---TGG	883
Db	153 AlaIysGlyTyProValGlnMetIeuIleIeuThrGlyIeuIeuPheGlyAlaIeuIeu	172
OY	884 AATTCGTTTTGGAAATGATTTTAAAGAGGGTATGCTTTTAACTTTAAATCCAAAAAG	943

Db 173 AsnAlaLeuLeuTyLeu-:::|||||||::: 186
Qy 944 CATGCATCAGCGATGTGCTCAAAATGCGTGGCAAAAGTTGTGCATAAACTACAGGAAGCT 1003
Db 187 -----MetAsnProIleAla-:::|||||||:::SerTyrLeuPheGly 196
Qy 1004 GGTCTTCTCTAGACATTCGGT-----ATCAACATTTGGATCTTCAGAA 1051
Db 197 GlyPheAlaSerAlaGluTyrGlnAspValMetIleIleSerLeuIleAlaSerValAla 216
Qy 1052 AAGATCAAAATCTATTTAGTGATAAAGCTTTAAGCTATATAAG-----CAGATAGA 1105
Db 217 llelleValLeuPheLeuMetGlnlysGlyIleIleLeuLeuGlnValGlyIleLeuLys 236
Qy 1106 GCCTCTCTCTAAATTAACGATCATGAGCTGGCGTTATTGTGGGATGTGTGTCAGAAAC 1165
Db 237 SerGlnSerLeuGlyLeuAsnValGlnGlnValThrTyrIleValLeuIleVal----- 254
Qy 1166 AGGCCTAGATTTCTCTACGGAACTCTAAACGAACGCAAAATTTGGTCAAGGTAAGC 1225
Db 255 -----AlaSerIleMetThrAlaVal--- 261
Qy 1226 AGCAAACTATCGAAGAAATGCGTTATCAGCGACCATCGGCTTTTAGGAGCTTTGGCA 1285
Db 262 -----ValValAlaTyrValGlyValIleGlyPheIleGly 273
Qy 1286 ATCATCTTGCTCTATGTGAGTTTGCGCTTTGAATGGCAATATGCTTTTCAGTCCGCTATGC 1345
Db 274 MetIleIleProGlnLeuIleargLysPheTyrTrpAgtYrGlnIleGly-----Leu 291
Qy 1346 GCTTAAATCATGACCTTTGGCTACCTGTGCGAGCTGTGTGTGTATAGCACATTC----- 1399
Db 292 GlnMetLeuLeuAsnIleIleIleGlyAlaValValMetIleIleAlaAspPheIleGly 311
Qy 1400 -----TTTTGAAGAAATTCATATAGATTGCAAGCATTTGGTCTTTAATGACTGTA 1453
Db 312 SerThrMetIleGlnProIleGlnIle-----ProValGlyIleValMetAlaLeu 328
Qy 1454 TTGGGCTATTCAATCAAACTACTTTGATCATTTTGTATGCTGTTGCTGGAAGTCCCAA 1513
Db 329 LeuGlyValProValLeuPheTyrIleLeuIle-----LysGln 341
Qy 1514 GCGAACCTG 1522
Db 342 ThrAsnIle 344

RESULT 39
US-08-118-270-73
; Sequence 73, Application US/08118270
; Patent No. 5508384
; GENERAL INFORMATION:
; APPLICANT: Murphy, Randall B.
; APPLICANT: Schuster, David I.
; TITLE OF INVENTION: POLYPEPTIDES OF G-COUPLED PROTEIN
; TITLE OF INVENTION: RECEPTORS, AND COMPOSITIONS AND METHODS THEREOF
; NUMBER OF SEQUENCES: 348
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK
; STREET: 419 Seventh Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/118,270
; FILING DATE: 09-SEP-1993
; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/943,236
; FILING DATE: 10-SEP-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Townsend, Kevin G.
; REGISTRATION NUMBER: 34,033
; REFERENCE/DOCKET NUMBER: MURPHY-2A
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; TELEX: 248633
; INFORMATION FOR SEQ ID NO: 73:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 334 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-118-270-73

Alignment Scores:
Pred. No.: 0.0789 Length: 334
Score: 92.50 Matches: 82
Percent Similarity: 33.50% Conservative: 54
Best Local Similarity: 20.20% Mismatches: 115
Query Match: 2.79% Indels: 155
DB: 1 Gaps: 22

US-09-868-987-1 (1-1864) x US-08-118-270-73 (1-334)

Qy 290 ATCTCAGCATGCTGCTGGCTGGCAATCTTATTGTTTGATGAGCGGTATATATAGATTT 349
Db 7 IleLeuIleCys---PheIleIleLeuGluAsnIlePheVal----- 20
Qy 350 GGAGCGCATCGCTCGGGAGCTTCTCTCTGAATCTTTGCTTATCTGCGGAGCT--- 406
Db 21 -----LeuLeuThrIleTrpLysThrLys 28
Qy 407 -----CTACAGTATTGGATGCGCCACTCACCTTTGTCTAGGACTCGCTGGG 451
Db 29 LysPheHisArgProMetTyrTyrPheIleGlyAsnIleAlaLeuSerAspLeuIleAla 48
Qy 452 ATTGTTCTTGCTATGGGATGGCGGTAGATGCAAAATGTTCTTGTATTGGAAGAATCCGA 511
Db 49 -----GlyValAlaTyrThrAlaAsnLeuLeu----- 58
Qy 512 GAGCAATTTTATTGCTCTCAAAGCTTAAATAATCTGTAGAAAAGGATATACCAAGSCT 571
Db 58 ----- 58
Qy 572 TTGGAGCCATTTTGTATTCTAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 631
Db 59 -----SerGlyAlaThrThrTyrLysLeuThrProAlaGlnTrpPhe 72
Qy 632 CTAGATACAGGCGCTTAAAGGTTGCTTTGACATTTGATTGATTGATTGATTGATTGATTGATT 691
Db 73 LeuArgGluGly-----SerMetPheValAlaLeuSerLeuSerValPheSerLeu 89
Qy 692 ATGTTTACGCTCTTTTCATGACTAAATTTTCTTCATGCTGTGTGATG-----AATAAG 745
Db 90 Leu-----AlaIleAlaIleGluArgTyrIleThrMetLeuLysMetLeuHisAsnGly 107
Qy 746 ACCCAACATACACAGTGTGCATATGATGAATAAGTTCGTGGGGATAAAGCATGATTTCTTGT 805
Db 108 SerAsnAsnPheArgLeuPheLeu-----Ile 117
Qy 806 AGAGGATGCAAAAACTTTGGCTGTTTCTGGAAGTGTTTTCTTTAGTGTGCTGCTGCT 865
Db 118 SerAlaCys-----TrpValIleSer-----LeuIleLeuGly-----Gly 129
Qy 866 CTCGGGTTTGGAGCCTCGAATTCGTTTGGGATGATTTTAAAGGAGGATGATGCTGCTTT 925
Db 130 LeuProIleMetGlyTrpAsnCysIle----- 138

```

QY 926 ACCTTAAATCCAAAAGACATGCATGACGATGTCCTCAATGCGTGGCAAGTTGTG 985
DB 138 ----- 138
QY 986 CATAACTACAGAGAGCGTCTTTCTTCTAGAGACTTCGATTCAACATTGGATCT 1045
DB 139 -----SerAlaLeuSerSer-----Cys 144
QY 1046 TCAGAAAAATCAAAATCTATTGATTAAGATTAAGCTTAAGCTTAAGCAGATACGA 1105
DB 145 SerThrValLeuProLeuTyrHisIleHisTyrIleLeu---PheCysThrLeuIleVal 163
QY 1106 GCCTCTCTCTAAATTAAGCATATGAGCTGGCGTTATTTGGGATT-----CTGTGC 1159
DB 164 PheThrLeuLeuLeuSerIleValIle---LeuTyrCysArgIleTyrSerLeuVal 182
QY 1160 AGAAACAGCGCTGATTTCTCTACGAAACTTAACGAAACGCAAAATTTTGGTCAAG 1219
DB 183 ArgThrArgSerArgArgLeu-----ThrPheArgLysAsn----- 194
QY 1220 GTACGACCAAACTATCGAAGAAATGCGTTATCAGCGACCATCGGCGTTTGAAGCT 1279
DB 195 -----IleSerLysAlaSerArgSerSerGluAsnValAlaLeuLeuLysThr 210
QY 1280 TTGGCATCATCTTG---CTTATGTAGAGTTGGCGTTTGAATGGCATATAGCTTTGAGT 1336
DB 211 ValIleIleValLeuSerValPheIleAlaCys-----TyrAlaProLeuPheIle 227
QY 1337 GCGGTATGCGGTTTAAATCATGACCTTTGGCTACCTGTGACCTTGTATATGACACAT 1396
DB 228 LeuLeuLeuLeuValIleGlyCysLeuValIleThrCysAspIleLeuPheArgAlaGlu 247
QY 1397 TTCTTTTGAAGAAATTCAAATAGATTTCGACCCATGCTGCTTAATGACTGTATG 1456
DB 248 TyrPheLeuVal-----IleAlaValIleAsnSerGlyThrAsnProIle 262
QY 1457 GCGTATTCATTAACAT 1474
DB 263 IleTyrThrLeuThrAsn 268

RESULT 40
PCT-US93-08528-73
Sequence 73: Application PC/TUS9308528
GENERAL INFORMATION:
APPLICANT: New York University
TITLE OF INVENTION: POLYPEPTIDES OF G-COUPLED PROTEIN
TITLE OF INVENTION: RECEPTORS, AND COMPOSITIONS AND METHODS THEREOF
NUMBER OF SEQUENCES: 348
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROWDY AND NEIMARK
STREET: 419 Seventh Street, N.W., Suite 300
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/08528
FILING DATE: 09-SEP-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/943,236
FILING DATE: 10-SEP-1992
ATTORNEY/AGENT INFORMATION:
NAME: Townsend, Kevin G.
REGISTRATION NUMBER: 34,033
REFERENCE/DOCKET NUMBER: MURPHY=2 PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528

```

```

; TELEX: 248633
; INFORMATION FOR SEQ ID NO: 73:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 334 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; PCT-US93-08528-73

Alignment Scores:
Pred. No.: 0.0789 Length: 334
Score: 92.50 Matches: 82
Percent Similarity: 33.50% Conservative: 54
Best Local Similarity: 20.20% Mismatches: 115
Query Match: 2.79% Indels: 155
DB: Gaps: 22

US-09-868-987-1 (1-1864) x PCT-US93-08528-73 (1-334)
QY 290 ATCTCAGCATGCTGTGGCTTGGCAATGCTTAATGTTTGAATGAGCGTAAATTAAGATT 349
DB 7 IleLeuIleCysCys-----PheIleIleGluAsnIlePheVal----- 20
QY 350 GAGGCGTATCGCTTCGGAGCGTCTTCTGAATCTTTTGGTAAATTCGGGACGCT--- 406
DB 21 -----LeuLeuThrIleTyrLys 28
QY 407 -----CTACAGTATTGATGCGCCACTACCTTGTACAGGACTCGCTGGG 451
DB 29 LysPheHisArgProMetCysTyrPheIleGlyAsnIleAlaLeuSerAspLeuIleAla 48
QY 452 ATGTCTTGTCTATGGGAGTGCCTGATGCAATGTTCTGTATTCGAAGAATCCGA 511
DB 49 -----GlyValAlaTyrThrAlaAsnLeuLeu----- 58
QY 512 GAGGAATTTTATTGTCCTCAAACTCTTAAGAAATCTGTAGAAAGATATACCAAGCT 571
DB 58 ----- 58
QY 572 TTGAGACCATTTTGTATCTTAACCTTACATGATGATGAGCTTCACACTCTTTCTTTC 631
DB 59 -----SerGlyAlaThrThrTyrLysLeuThrProIleGlnTyrPhe 72
QY 632 CTAGATACAGGCGCTAATTAAGGCTTGTCTGACATGATTTTGAAGATTTCCTTCA 631
DB 73 LeuArgGluGly-----SerMetPheValAlaLeuSerLeuSerValPheSerLeu 89
QY 692 ATGTTTACGCTCTTTTCATGACTAATTTTCTTCAATGCTGTGATG-----AATAAG 745
DB 90 Leu-----AlaIleAlaIleGluArgTyrIleThrMetLeuLysMetLeuHisAsnGly 107
QY 746 ACCCAACATACACAGTTGATGATGATGATGATGCTGTGGGATTAAGCATGATTTCTTG 805
DB 108 SerAsnAsnAsnLeuPheLeu-----Ile 117
QY 806 AGAGATGCAAAAAAATCTGGGCTGTTTCTGAGAGTTTCTTTAGGCTCGTGTCT 865
DB 118 SerAlaCys-----TyrValIleSer-----LeuIleGluGly-----Gly 129
QY 866 CTGCGGTTTGAAGCTGGAATTCCTGTTTGGGAATGATTTTAAAGAGGATATGCTTT 925
DB 130 LeuProIleMetGlyTyrPheAsnCysIle----- 138
QY 926 ACCTTAAATCCAAAAGACATGCATGACGATGTCCTCAATGCGTGGCAAGTTGTG 985
DB 138 ----- 138
QY 986 CATAACTACAGAGAGCGTCTTTCTTCTAGAGACTTCGATTCAACATTGGATCT 1045
DB 139 -----SerAlaLeuSerSer-----Cys 144
QY 1046 TCAGAAAAATCAAAATCTATTGATTAAGATTAAGCTTAAGCTTAAGCAGATACGA 1105

```

```
Db 145 SerThrValLeuProLeuTyrHisLysHisTyrIleLeu---PheCysThrLeuIleVal 163
Qy 1106 GCCTCTCTCCTAAATACCATCATGAGTCGGCTGTTATTGGGATT-----GTTGTC 1159
Db 164 PheThrLeuLeuLeuSerIleValIle---LeuTyrCysArgIleTyrSerLeuVal 182
Qy 1160 AGAACAGCGCTAGATTCTCTACGGAAACTCTAAACGAAACGCAAAATTTTGGTCAAAG 1219
Db 183 ArgThrArgSerArgArgLeu-----ThrPheArgLysAsn----- 194
Qy 1220 GTAACGACAAATCTCAAGAAATGCGTTATACGGGACCATCGGCTTTTAGGAGCT 1279
Db 195 -----IleSerLysAlaSerArgSerGluAsnValAlaLeuLeuLysThr 210
Qy 1280 TTGGCAATCATCTTG---CTCTATGTAGTTTTCGCTTTTGAATGGCAATATGCTTTTCACT 1336
Db 211 VallIleIleValLeuSerValPheIleAlaCys-----TipAlaProLeuPheIle 227
Qy 1337 GCGGTATGCGCTTAAATCATGACCTTTTGGTACTGTCAGTCTTGTATTATACACAT 1396
Db 228 LeuLeuLeuLeuAspValGlyCysLysValLysThrCysAspIleLeuPheArgAlaGlu 247
Qy 1397 TCTCTTTTGAAGAAATTCARATAGATTGCAAGCATTTGGTCTTTAATGACTGTATTG 1456
Db 248 TyrPheLeuVal-----IleAlaValIleAsnSerGlyThrAsnProIle 262
Qy 1457 GGGTATTTCATTAACAAT 1474
Db 263 IleTyrThrLeuThrAsn 268

RESULT 41
US-09-387-699-2
; Sequence 2, Application US/09387699
; Patent No. 6221660
; GENERAL INFORMATION:
; APPLICANT: Bonini, James A.
; APPLICANT: Borowsky, Beth E.
; APPLICANT: Adham, Nika
; TITLE OF INVENTION: DNA Encoding SNORP25 Receptor
; FILE REFERENCE: 56095-A
; CURRENT APPLICATION NUMBER: US/09/387,699
; PRIORITY FILING DATE: 1999-08-13
; EARLIER APPLICATION NUMBER: 09/255,376
; EARLIER FILING DATE: 1999-02-22
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn Ver. 2.0 - beta
; SEQ ID NO 2
; LENGTH: 335
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-387-699-2

Alignment Scores:
Pred. No.: 0.0903 Length: 335
Score: 92.00 Matches: 65
Percent Similarity: 37.46% Conservative: 59
Best Local Similarity: 19.64% Mismatches: 139
Query Match: 2.78% Indels: 68
DB: 4 Gaps: 15

US-09-868-987-1 (1-1864) x US-09-387-699-2 (1-335)

Qy 434 TTGTGAGGACTCGCTGGGATCTCTTCTGATGGGATGGCGGTAGATGCATGCAATGCTTCT 493
Db 11 LeuAlaValLeuAlaSerLeuIleIleAlaThrAsnThrLeuValAlaValAlaValLeu 30
Qy 494 GTATTGGAAGAATCCGAGGAATTTTATTGTCTCAAGTCTTAAATAATCTAGAA 553
Db 31 LeuLeuIleHisLysAsnAspGlyValSerLeuCysPheThrLeuAsnLeuAlaValAla 50
Qy 554 AAAGATATACCAAGCTTTTGGAGCCATTTTGTGATCTTAACCTGACTACAGTATGGCC 613
```

```
Db 51 -----AspThrLeuIleGlyValAlaIleSerGlyLeuLeuThrAspGlnLeuSer 67
Qy 614 TCAGCAGCTCTCTTCTCTAGATACAGGGCTATTAAAGGGTTTGTCTTTTCACATTGATT 673
Db 68 Ser-----ProSerArgProThrGlnLysThrLeuCysSerLeuArg 81
Qy 674 TTAGAATTTTCTCTCAATGTTTACGGCTCTTTTCATGATCAATAATTTTCTTCATGCTG 733
Db 82 MetAlaPheValThrSerSerAlaAlaSerValLeuThr-----ValMet 97
Qy 734 TGGATGAATAGACCCACATACACAGTTGCATATGATGAATAGTTCGTGGGATAAAG 793
Db 98 LeuIleThrPheAspArgTyrLeuAlaIleLysGlnProPheArgTyrLeuLysIleMet 117
Qy 794 CATGATTTCTTGAGAGGATGC-----AAAAAACTTTGGGCTGTTCTTGGGAAGTGTTTT 847
Db 118 SerGlyPheValAlaGlyAlaCysIleAlaGlyLeuTrpLeuValSer-----Tyr 134
Qy 848 CTTTATGTTGCGTCTCTCGGGTTTGGAGCCTGGGAATCCGTTTGGGAATGGAATTTT 907
Db 135 LeuIleGlyPheLeuProLeuGlyIleProMetPheGlnGlnThrAla-----Tyr 151
Qy 908 AAAGGAGGTATGCTTT-----ACCTTTAATCCAAAAGAGCAT----- 946
Db 152 LysGlyGlnCysSerPhePheAlaValPheHisPro-----HisPheValLeuThrLeu 169
Qy 947 -----GGCATCAGC 955
Db 170 SerCysValGlyPhePheProAlaMetLeuLeuPheValPhePheTyrCysAspMetLeu 189
Qy 956 GATGTTGCTCAATGCGTGGCAAGTTGTGCATAAACTACAGGAAGCTGCTCTTCTTCTT 1015
Db 190 LysIleAlaSerMetHisSerGlnGlnIleArgLysMetGluHisAlaGlyAlaMetAla 209
Qy 1016 AGAGACTTCCGTATT-----CAACATTTGGATCTTCAGAAAAGATCAAAATC 1063
Db 210 GlyGlyTyrArgSerProArgThrProSerAspPheLysAlaLeuArgThrValSerVal 229
Qy 1064 TATTTAGTGATAAAGCTTAAAGTATATAGCTATACAGCAGATCAGGCTCTCTCTAAAATTA 1123
Db 230 LeuIleGlySerPheAlaLeuSerTrpThrProPheLeuIleThrGlyIleValGlnVal 249
Qy 1124 ACGATCATGAGCTGGCGTTATTGTGGGATTGTTGTGCAGAAAACAGGCTTAGATTTCTTAC 1183
Db 250 AlaCysGlnGluCysHisLeuTyrLeuValLeu-----GluArgTyrLeuTrp 265
Qy 1184 -----GGAAACTCTAAAGAAACGCA-----AAATTTGGTCAAG--- 1219
Db 266 LeuLeuGlyValGlyAsnSerLeuLeuAsnProLeuIleTyrAlaTyrTrpGlnLysGlu 285
Qy 1220 GTAAGCAGCAAACTATCGAAGAAATGCGTTATCAGGCGACCATCGGCTTTTAGGAGCT 1279
Db 286 ValArgLeuGlnLeu-----TyrHisMetAlaLeuGlyValLysLysVal 300
Qy 1280 TTGCAATCATCTTCTCTCTATGTGAGTTTGGC 1312
Db 301 LeuThrSerPheLeuLeuPheLeuSerAlaArg 311

RESULT 42
US-09-641-259B-2
; Sequence 2, Application US/09641259B
; Patent No. 6468756
; GENERAL INFORMATION:
; APPLICANT: Bonini, James A
; APPLICANT: Borowsky, Beth E
; APPLICANT: Adham, Nika
; APPLICANT: Boyle, No. 64687561
; APPLICANT: Thompson, Thelma O.
; TITLE OF INVENTION: DNA Encoding SNORP25 Receptor
; FILE REFERENCE: 1995/56095-B/JPW/ADM
; CURRENT APPLICATION NUMBER: US/09/641,259B
; CURRENT FILING DATE: 2002-03-12
```

; PRIOR APPLICATION NUMBER: PCT/US00/04413
 ; PRIOR FILING DATE: 2000-02-22
 ; PRIOR APPLICATION NUMBER: US 09/387,699
 ; PRIOR FILING DATE: 1999-08-13
 ; PRIOR APPLICATION NUMBER: US 09/255,376
 ; PRIOR FILING DATE: 1999-02-22
 ; NUMBER OF SEQ ID NOS: 35
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 2
 ; LENGTH: 335
 ; TYPE: PR1
 ; ORGANISM: Homo sapiens
 ; US-09-641-2598-2

Alignment Scores:
 Pred. No.: 0.0903 Length: 335
 Score: 92.00 Matches: 65
 Percent Similarity: 37.46% Conservative: 59
 Best Local Similarity: 19.64% Mismatches: 139
 Query Match: 2.78% Indels: 68
 DB: Gaps: 15

US-09-868-987-1 (1-1864) X US-09-641-2598-2 (1-335)

QY 434 TTGTCAGAGACTCGCTGGATGTTCTGCTATGGAGCGCCGATGCAATGTTCTT 493
 Db 11 LeuValValLeuValSerLeuLeuLeuValLeuValLeuValLeuValLeu 30
 QY 494 GTATTCAGAAAGATCCGAGAGATTTTATTTGCTCAAGCTTAAATAATCTGTGAA 553
 Db 31 LeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 50
 QY 554 AAAGATATACCAAGCTTTTGAGCCATTTTGAATCTTAACTTACATGATGAGCC 613
 Db 51 -----AspThrLeuLeuLeuValAlaLeuSerGlyLeuLeuThrAspGlnLeuSer 67
 QY 614 TCAGCACTTCTTTCTTCTGATACAGAGGCTATTAAGGTTTGCTTACATGATG 673
 Db 68 Ser-----ProSerArgProThrGlnGlySerThrLeuCySerLeuArg 81
 QY 674 TTAGGAATTTCTCTCAATGTTTAAAGGCTTTTCAATGACTTAAATTTTCTTCATGCTG 733
 Db 82 MetAlaPheValInrSerSerAlaAlaLeuValLeuInr-----ValMet 97
 QY 734 TGGATGAATAGCAACCAATACACAGTTGCATATGATGAATAGTTCGTGGGATTAAG 793
 Db 98 LeuLeuThrPheAspArgTyrLeuAlaLeuLeuGlnProPheArgTyrLeuLeuIleMet 117
 QY 794 CATGATTTCTTGAGAGATGC-----AAAAAATCTTGGGCTGTTTCTGGAAGTCTTTT 847
 Db 118 SerGlyPheValAlaGlyAlaCysIleAlaGlyLeuThrLeuValSer-----Tyr 134
 QY 848 CTTTATGAGTTCGCTGCTGCTGGGTTTGAGCCGATTCGTTTGGGAATGATTTT 907
 Db 135 LeuLeuGlyPheLeuProLeuGlyIleProMetPheGlnInrAla-----Tyr 151
 QY 908 AAAGAGAGGATATGCTTT-----ACCTTTATCCAAAAGAGCAT----- 946
 Db 152 LysGlyGlnCysSerPhePheAlaValPheHisPro-----HisPheValLeuThrLeu 169
 QY 947 -----GGCATCAGC 955
 Db 170 SerCysValGlyPhePhePheAlaMetLeuLeuPheValPhePheTyrCysAspMetLeu 189
 QY 956 GATTTCTGCAAAAGCGG30AAAGTTGTGATTAACACAGAAAGCTGCTTTCTTCT 1015
 Db 190 LysLeuAlaSerMetHisSerGlnGlnIleArgLysMetGluHisAlaGlyAlaMetAla 209
 QY 1016 AGAGATCCGCTATTT-----CAAAATTTGGATCTTCAAGAAAGATCAAAATC 1063
 Db 210 GlyGlyTyrArgSerProArgThrProSerAspPheLysAlaLeuAlaGlnThrValSerVal 229
 QY 1064 TATTTAGTATTAAGCTTTAAGCTATATCAAGAGATACGAGCGCTCTCTCTAAATTA 1123

Db 230 LeuIleGlySerPheAlaLeuSerTrpThrProPheLeuIleThrGlyIleValGlnVal 249
 QY 1124 ACGATCATGAGCTGGGCTTATGTTGGATGTTGTCAGAAAGCGCTGATTTCTTCTAC 1183
 Db 250 AlaCysGlnGluCysHisLeuTyrLeuValLeu-----GlnArgTyrLeuTyrP 265
 QY 1184 -----GGAACTCTAAACGAAAGCGCA-----AAATTTGGTCAAG--- 1219
 Db 266 LeuLeuGlyValGlyAsnSerLeuLeuAsnProLeuIleTyrAlaTyrTrpGlnLysGln 285
 QY 1220 GTAAGCAGCAACTATTCGAAAGAAATGCTTATCAGCGCAGCATCGGCTTTAGAGAGCT 1279
 Db 286 ValArgLeuGlnLeu-----TyrHisMetAlaLeuGlyValLysVal 300
 QY 1280 TTGGCAATCATCTTGCTCTATGATGATGCGC 1312
 Db 301 LeuThrSerPheLeuLeuPheLeuSerAlaArg 311

RESULT 43
 US-08-118-270-64
 ; Sequence 64, Application US/08118270
 ; Patent No. 5508384

GENERAL INFORMATION:

; APPLICANT: Murphy, Randall B.
 ; TITLE OF INVENTION: POLYPEPTIDES OF G-COUPLED PROTEIN
 ; TITLE OF INVENTION: RECEPTORS, AND COMPOSITIONS AND METHODS THEREOF
 ; NUMBER OF SEQUENCES: 348
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSER: BROWDY AND NEIMARK
 ; STREET: 419 Seventh Street, N.W., Suite 300
 ; CITY: Washington
 ; STATE: D.C.
 ; COUNTRY: USA
 ; ZIP: 20004

COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/118,270
 ; FILING DATE: 09-SEP-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/943,236

FILING DATE: 10-SEP-1992

ATTORNEY/AGENT INFORMATION:

NAME: Townsend, Kevin G.

REGISTRATION NUMBER: 34,033

REFERENCE/DOCKET NUMBER: MURPHY=2A

TELECOMMUNICATION INFORMATION:

TELEPHONE: 202-628-5197

TELEFAX: 202-737-3528

TELEX: 248633

INFORMATION FOR SEQ ID NO: 64:

SEQUENCE CHARACTERISTICS:

LENGTH: 269 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

US-08-118-270-64

Alignment Scores:
 Pred. No.: 0.105 Length: 269
 Score: 91.00 Matches: 44
 Percent Similarity: 37.82% Conservative: 29
 Best Local Similarity: 22.80% Mismatches: 70
 Query Match: 2.75% Indels: 50
 DB: Gaps: 8

US-09-868-987-1 (1-1864) X US-08-118-270-64 (1-269)

```

; TELEFAX: 202-737-3528
; TELEX: 248633
; INFORMATION FOR SEQ ID NO: 64:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 269 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; PCT-US93-08528-64

Alignment Scores:
Pred. No.: 0.105 Length: 269
Score: 91.00 Matches: 44
Percent Similarity: 37.82% Conservative: 29
Best Local Similarity: 22.80% Mismatches: 70
Query Match: 2.75% Indels: 50
DB: 8

US-09-868-987-1 (1-1864) x PCT-US93-08528-64 (1-269)

QY 1223 AGCAGCAAACTATCGAAGAAAATCGTTATCAGCGGACC---ATCGGGCTTTTAGGAGCT 1279
Db 50 SerLeuLysLeuLeuGlnAsnMetArgSerGlnAspThrSerIleProTyrGlyGlyCys 69

QY 1280 TTGCA-----ATCATCTTGCTCTATGTGAGTTTG 1309
Db 70 LeuAlaGlnThrTyrPhePheMetValPheGlyAspLeuSerPheLeuLeuValAlaMet 89

QY 1310 CGCTTTGAATGGCAATATGCTTCAGTCGCGTATGCGCTTTAAATTCAT----- 1357
Db 90 AlaTyrAsp-----ArgTyrValAlaIleCysPheLeuProHisTyrThrSerIle 106

QY 1358 -----GACCTTTGGCTACTGTCAGTCTTTCAGTCGCGTATGCGCTTTAAATTCAT----- 1357
Db 107 MetSerProLysLeuLeuValLeuLeuValLeuLeuValLeuLeuValLeuLeuVal 120

QY 1412 ATTCAATAGATTTTGAAGCCATTTGCTTTAATGCTGTTTGGGTATTCATTAAAC 1471
Db 121 -----MetLeuThrSerHisMetMetThrLeuLeuLeuAlaAlaArgLeuSer 136

QY 1472 AATATTGATCATTTTGTATCGTATTCGTAAGATCGCAAGCACTGTTTACCCCT 1531
Db 137 -----PheCysGluAsnAsnTrpLeuAsnPhePheCysAsp 148

QY 1532 ATGCATGTTTGTAGTTAATGATGCTTCCTTCAAAAGACGTTTCAGCCGACGGTAATGACAACA 1591
Db 149 LeuPheValLeuLeuLysIleAlaCysSerAspThrTyrIleAsnGluLeuPheIleMet 168

QY 1592 GCTACAACTCTATCAGTTTGTGTTAATGCTTTTGTATTA----- 1630
Db 169 SerThrLeuLeuIleIleIleProPhePheLeuIleValMetSerTyrAlaLysValPro 188

QY 1631 -----GCGGCTCTCTGCTCTTTAAATTTT-----GCATTTATTATGACCATAGG 1675
Db 189 SerThrGlnGlyIleCysLysValPheSerThrCysGlySerHisLeuSerValValSer 208

QY 1676 ATTCTTCTAGGAAGCTTTATCGTCTCTTTATATTCACCA 1714
Db 209 LeuPheTyrGlyThrIleIleGlyLeuTyrLeuCysPro 221

RESULT 44
PCT-US93-08528-64
; Sequence 64, Application PC/TUS9308528
; GENERAL INFORMATION:
; APPLICANT: New York University
; TITLE OF INVENTION: POLYPEPTIDES OF G-COUPLED PROTEIN
; TITLE OF INVENTION: RECEPTORS, AND COMPOSITIONS AND METHODS THEREOF
; NUMBER OF SEQUENCES: 348
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK
; STREET: 419 Seventh Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/08528
; FILING DATE: 09-SEP-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/943,236
; FILING DATE: 10-SEP-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Townsend, Kevin G.
; REGISTRATION NUMBER: 34,033
; REFERENCE/DOCKET NUMBER: MURPHY=2 PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197

```

```

; TELEFAX: 202-737-3528
; TELEX: 248633
; INFORMATION FOR SEQ ID NO: 64:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 269 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; PCT-US93-08528-64

Alignment Scores:
Pred. No.: 0.105 Length: 269
Score: 91.00 Matches: 44
Percent Similarity: 37.82% Conservative: 29
Best Local Similarity: 22.80% Mismatches: 70
Query Match: 2.75% Indels: 50
DB: 8

US-09-868-987-1 (1-1864) x PCT-US93-08528-64 (1-269)

QY 1223 AGCAGCAAACTATCGAAGAAAATCGTTATCAGCGGACC---ATCGGGCTTTTAGGAGCT 1279
Db 50 SerLeuLysLeuLeuGlnAsnMetArgSerGlnAspThrSerIleProTyrGlyGlyCys 69

QY 1280 TTGCA-----ATCATCTTGCTCTATGTGAGTTTG 1309
Db 70 LeuAlaGlnThrTyrPhePheMetValPheGlyAspLeuSerPheLeuLeuValAlaMet 89

QY 1310 CGCTTTGAATGGCAATATGCTTCAGTCGCGTATGCGCTTTAAATTCAT----- 1357
Db 90 AlaTyrAsp-----ArgTyrValAlaIleCysPheLeuProHisTyrThrSerIle 106

QY 1358 -----GACCTTTGGCTACTGTCAGTCTTTCAGTCGCGTATGCGCTTTAAATTCAT----- 1357
Db 107 MetSerProLysLeuLeuValLeuLeuValLeuLeuValLeuLeuValLeuLeuVal 120

QY 1412 ATTCAATAGATTTTGAAGCCATTTGCTTTAATGCTGTTTGGGTATTCATTAAAC 1471
Db 121 -----MetLeuThrSerHisMetMetThrLeuLeuLeuAlaAlaArgLeuSer 136

QY 1472 AATATTGATCATTTTGTATCGTATTCGTAAGATCGCAAGCACTGTTTACCCCT 1531
Db 137 -----PheCysGluAsnAsnTrpLeuAsnPhePheCysAsp 148

QY 1532 ATGCATGTTTGTAGTTAATGATGCTTCCTTCAAAAGACGTTTCAGCCGACGGTAATGACAACA 1591
Db 149 LeuPheValLeuLeuLysIleAlaCysSerAspThrTyrIleAsnGluLeuPheIleMet 168

QY 1592 GCTACAACTCTATCAGTTTGTGTTAATGCTTTTGTATTA----- 1630
Db 169 SerThrLeuLeuIleIleIleProPhePheLeuIleValMetSerTyrAlaLysValPro 188

QY 1631 -----GCGGCTCTCTGCTCTTTAAATTTT-----GCATTTATTATGACCATAGG 1675
Db 189 SerThrGlnGlyIleCysLysValPheSerThrCysGlySerHisLeuSerValValSer 208

QY 1676 ATTCTTCTAGGAAGCTTTATCGTCTCTTTATATTCACCA 1714
Db 209 LeuPheTyrGlyThrIleIleGlyLeuTyrLeuCysPro 221

RESULT 45
US-08-846-762-92
; Sequence 92, Application US/08846762A
; Patent No. 5994072
; GENERAL INFORMATION:
; APPLICANT: Lam, Joseph S.
; APPLICANT: Burrows, Lori
; APPLICANT: Charter, Deborah
; APPLICANT: de Kievit, Teresa
; TITLE OF INVENTION: No. 5994072el Proteins Involved in the Synthesis and Assembly
; TITLE OF INVENTION: of O-Antigen in Pseudomonas Aeruginosa
; FILE REFERENCE: 6580-089

```

CURRENT APPLICATION NUMBER: US/08/846,762A
CURRENT FILING DATE: 1997-04-30
NUMBER OF SEQ ID NOS: 100
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 92
LENGTH: 341
TYPE: PRF
ORGANISM: Pseudomonas aeruginosa
US-08-846-762-92

Alignment Scores:
Pred. No.: 0.119 Length: 341
Score: 91.00 Matches: 47
Percent Similarity: 35.47% Conservative: 25
Best Local Similarity: 23.15% Mismatches: 65
Query Match: 2.75% Indels: 66
DB: 2 Gaps: 9

US-09-868-987-1 (1-1864) x US-08-846-762-92 (1-341)

```
QY 344 AGATTGGAGCGCTCATCGCTTCGGAGCTTCTTGAATCTTTGCTTATCTGGGCA 403
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 47 ArgGlyGlyGlyValAlaIleValIleSerPheLeuIleGlyIleIleLeuPheTyr--- 65
QY 404 GCTCTACAGTATTGGATGCGCACTCACCTTTCAGAGCTCGCTGGGATGTTCTTGC 463
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 66 -----PheLeuGlyTyrLeuProIleLeuSerVal 75
QY 464 ATGGGGATGGCGCTGATGCAATGCTTCTTATTCGAAAGATCCGAGAGAAATTTTGA 523
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 76 ValGlyLeuIleValSerGlyValIleAlaLeuValGlyPheTyrPasp----- 92
QY 524 TTGCTCAAGCTTTAAAAATCTGTAGAAAAGATATACCAAGCTTTTGGAGCCATT 583
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 93 -----AspHisGlyHisIleAlaIleAlaArgTyrPargLeu 103
QY 584 TTGATTCTACTGACTAGTATGCTATGCGCTCAGACCTTTTCTTCTAGATACAGGG 643
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 104 -----LeuAlaHisPheSerAlaAlaIlePheLeuPheCysPheGlyGly 119
QY 644 -----CCATATTAAGGCTTTGCTTTCGATTCATTGATTTTGAATTTCTCTTCA 691
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 120 PheProValLeuAsnValSerGlyPheIleIleGlu-----LeuGlyIlePheGlySer 137
QY 692 ATGTTTACGGCTTTTTCATGATCAATAATTTTTCATGCTGTGATGATTAAGACCCA 751
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 138 LeuPheGlyLeuLeu-----PheLeuValTyrPmet----- 147
QY 752 CATACACAGTTGCATATGATGATTAAGTTGCGGGATTAAGCATGATTTCTTGAGAGA 811
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 148 -----LeuAsnLeuTyrAsnPheMetAspGlyIle-----AspGlyLeuAlaSer 162
QY 812 TGCAAAAAAATTTGGGCTGTTTCTGGAAGTGT----- 844
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 163 AlaGluAlaValThrAlaCysIleGlyMetIleAlaIleTyrTyrIleSerGlyAspHis 182
QY 845 -----TTTCTTTAGGTTGCGTTGCTCTCGGGTTTGA 877
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 183 IleGluLeuAsnSerPheLeuValLeuTyrPheLeuAlaCysThrValLeuGlyPheLeu 202
QY 878 GCCTGGAAT 886
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 203 LeuTyrPasn 205
```

Search completed: December 17, 2002, 12:41:21
Job time : 81.5 secs

Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	44.3	13.4	472	1	SECD_SVNY3	Q55610 synecocyst
2	43.1	13.0	503	1	SECD_HELPY	Q526074 helicobacte
3	42.9	13.0	526	1	SECD_HELPJ	Q92166 helicobacte
4	42.7	12.9	501	1	SECD_AQUAE	Q67102 aquifex seo
5	42.6	12.9	583	1	SECD_TREPA	Q83425 treponema p
6	418.5	12.6	554	1	SECD_RHOCA	Q33517 rhodobacter
7	408.5	12.3	616	1	SECD_HAEIN	P44591 haemophilus
8	407.5	12.3	518	1	SECD_BRUAB	Q92986 brucella ab
9	398.5	12.0	518	1	SECD_RICPR	Q92986 brucella ab
10	364	11.0	323	1	SECF_ECOLI	Q92986 brucella ab
11	361	10.9	333	1	SECF_RHOCA	P19674 escherichia
12	360	10.9	324	1	SECF_RHOSH	Q33518 rhodobacter
13	359.5	10.9	615	1	SECD_ECOLI	Q33568 rhodobacter
14	355.5	10.7	615	1	SECD_SALCH	P19673 escherichia
15	351	10.6	323	1	SECF_HELPY	Q92166 helicobacte
16	350	10.6	323	1	SECF_HELPY	Q526073 helicobacte
17	342	10.3	311	1	SECF_RICPR	Q92986 brucella ab
18	342	10.3	325	1	SECF_HAEIN	Q92986 brucella ab

DR EMBL; D64000; BAA10118.1; -.
DR InterPro; IPR001036; AcRflvIn res.
DR InterPro; IPR003335; SecD_SecF.
DR Pfam; PF02355; SecD_SecF; 1.
DR PRINTS; PR00702; ACRIFLVINRNP.
DR TIGRfams; TIGR00916; 2A0604A01.1.

	TIGRfams; TIGR01129; secD; 1.
KM	Protein transposon; Translocation; Transmembrane; Membrane;
FT	Complete proteome.
FT	TRANSMEM 7 27 POTENTIAL.
FT	TRANSMEM 298 318 POTENTIAL.
FT	TRANSMEM 335 355 POTENTIAL.
FT	TRANSMEM 411 431 POTENTIAL.
FT	TRANSMEM 432 452 POTENTIAL.
SO	SEQUENCE 472 AA; 50468 MW; D246FEF2DCB263CA CRC64;
 Alignment Scores:	
Pred. NO.:	4.99e-28 Length: 472
Score:	443.00 Matches: 89
Percent Similarity:	62.39% Conservative: 57
Best Local Similarity:	38.03% Mismatches: 82
Query Match:	13.38% Indels: 6
DB:	Gaps: 1
 US-03-8668-987-1 (1-1864) x SHCD_SYNY3 (1-472)	
OY	50 CAATATTTCGAAACCGTGATGGCCGTGCTGAGTGAATGACGGTTAATGTCAGC 109
Db	213 GlnAlaValAlaGlyThrCylArgSerLeuGlyValPheIleAspAsnAspIleSer 232
OY	110 AGCCGTATTTTAAGCTGCCATTGAAAAT-----CATGCAGAGTGC 151
Db	223 AlaProvalValGlyValGluPheAlaasnThrGlyIleThrGlyGlyAlaIle 255
OY	152 TCAGGAAATTTACCACCGTAGAGTGAAGCAACTGCGCTCAGATTMAATCGAGCG 211
Db	253 ThrGlyAsnPhetheThrIleAspThrAlaasnSpleuAlaValGlnLeuArgGlySer 272
OY	212 ATGCTCTTTGTTCCGAGCTTCTCAGTGAAGACAGCATCTCTTGATCTTGGAAAAA 271
Db	273 LeuProPheProValGlyValGluAsnThrThraValGlyAlaThrLeuGlyGlnGlu 292
OY	272 CAAATGACAAGAAGATTATCTCACGATCGTGGCTGGCAATGCTTAATGTTGATG 331
Db	293 SerIleArgArgSerLeuValAlaGlyPheValGlyLeuValIleuValLeuValPheMet 312
OY	332 AGCGTATATTATAGATTGAGAGCGTCATCCGCTTCGGAGCGTTCCTGGAATCTTTG 391
Db	313 AlaValIlyrTyArgLeuProGlyIleValAlaAspIleSerLeuMetIleTyAlaVal 332
OY	392 CTATATCGGCGAGCTCTACAGTATTGAGATGCGCACTCACCTTGTGACAGCTCGTGG 451
Db	333 LeuThrIleuAlaAlaPheAlaLeuValAlaGlyAlThrLeuThrLeuProGlyIleAlaGly 352
OY	452 ATTGTTCTTTCGTAATGGGATGCGCCGTATAGCAAAATGTTGTAATCGAAGAATCGA 511
Db	353 PheIleLeuSerIleGlyMetAlaValAspAlaasnValIleuIlePheGlnArgThrArg 372
OY	512 GAGGAATTTTTATTGTCACAAAGTCTTAAAAATCTGAGAAAAAGATATACCAAGGCT 571
Db	373 GluGluLeuArgAlaGlyrAsnthrLeuTyrrArgSerValGlnAlaGlyPhePheArgAla 392
OY	572 TTGGAGCCATTTTGATTTCACTTAATGACATAAGATTGGCTGACACTTTCTTTCTTC 631
Db	393 PheSerSerIleLeuAspSerAsnValIthrThrIleuIleAlaCysAlaAlaIleuPheTrp 412
OY	632 CTAGATACAGGCGCCATTTPAAGGGTTTGCTTGACATTGATTTAGAAATTTCTCTCA 691
Db	413 PheGlySerGlyLeuValIleGlyPheAlaIleuThrIleuAlaIleGlyValMetValSer 433
OY	692 ATGTTACGGCTCTTTTCATGACTAAATTTTCTTCATGCTG 733
Db	433 LeuPheThrAlaLeuThrCysSerArgThrIleuLeuVal 446
 RESULT 2	
SEC2_HELPY STANDARD; PRT; 503 AA.	
ID	_SEC2_HELPY
AC	OZ6074;
DT	30-MAY-2000(Rel. 39, Created)

```

DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Protein-export membrane protein secD.
GN secD OR HP1550.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxId=210;
[1]
RA SEQUENCE FROM N.A.A.
RA STRAIN=26695 / ATCC 700392;
RA MEDLINE=97394467; PubMed=9252185;
RA Tomb J.-F., White O., Kariyaage A.R., Clayton R.A., Sutton G.G.,
RA Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
RA Nelson K., Quackenbush J., Zhou L., Kittnes E.F., Peterson S.,
RA Letunsky B., Richardson D., Dodson R., Khalak H.G., Glodek A.,
RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
RA Cotton M.D., Weidman J.M., Fujii C., Bowman C., Wathey L., Wallin E.,
RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
RA Venter J.C.;
RA "the complete genome sequence of the gastric pathogen Helicobacter
RA pylori."
RA Nature 388:539-547(1997).
RL -1- FUNCTION: INVOLVED IN PROTEIN EXPORT (BY SIMILARITY).
CC -1- SUBUNIT: PART OF THE PROKARYOTIC PROTEIN TRANSLLOCATION APPARATUS
CC WHICH COMPRISE SECY, SECB, SECD, SEEE, SECF, SECG AND SECY
CC (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).
CC -1- SIMILARITY: BELONGS TO THE SECD/SECY FAMILY. SECD FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE000652; AAD08588.1; -.
DR TIGR; HP1550; -.
DR InterPro; IPR001036; AcctFlvln.res.
DR InterPro; IPR00335; SecD.SecF.
DR Pfam; PF02355; SecD.SecF_1.
DR PRINTS; PR00702; ACRIPLAVINRP.
DR TIGRFAMs; TIGR00916; ZA0604s01; 1.
DR TIGRFAMs; TIGR01129; secD; 1.
KW Protein transport; Translocation; Transmembrane; Membrane;
KW Complete proteome.
FT FT TRANSMEM 199 219 POTENTIAL.
FT TRANSMEM 334 354 POTENTIAL.
FT TRANSMEM 357 377 POTENTIAL.
FT TRANSMEM 383 403 POTENTIAL.
FT TRANSMEM 456 476 POTENTIAL.
SQ SEQUENCE 503 AA; 54247 MW; 8541C291CA317086 CRC64;

Alignment Scores:
Pred. No.: 4.66e-27 Length: 503
Score: 431.00 Matches: 99
Percent Similarity: 60.83% Conservative: 47
Best Local Similarity: 41.25% Mismatches: 92
Query Match: 13.01% Indels: 2
Gaps: 2

US-09-868-987-1 (1-1864) x SECD_HELPY (1-503)
QY 47 GGACAAATTTCTC---GCAACCGTGAGGCGTATGAGCTGATGATGACGGTATATG 103
DB 232 GYAPSPneSeGgYAlaSnVaIcYgYsArGMeTAlaIleValLeuAspSnlySval 271
QY 104 GTACACACCCCTATTATTAACGTCCTCATTG--AAAAATCATGCGACAGTCTCAGAGAAA 160
DB 272 TYSrEtAlaPrOValIleArlgYAlaRglleGylgYglYseGylgYlnIleSerIyAen 291

```

```
QY 161 TTTACCACCGTGAAGTCAGCAAACTGCCTCAGATTTAAATCTGAGCGATCTCTTT 220
    |||...|||...|||...|||...|||...|||...|||...|||...|||...
Db 292 PheSerValAlaGlnAlaSerAspLeuAlaLeuAlaLeuArgSerGlyAlaMetSerAla 311
QY 221 GTTCCCGAGGTTCTCAGTCAGAGACGATCTCTCTGATCTTTGGGAAAAACAATGTACA 280
    |||...|||...|||...|||...|||...|||...|||...|||...|||...
Db 312 ProIleGlnValLeuGluLysArgIleGlyProSerLeuGlyLysAspSerValLys 331
QY 281 CAAGGCATTATCTCAGCATGCTGTGGCTGGCAATGCTTATTTGATGAGCGTATAT 340
    |||...|||...|||...|||...|||...|||...|||...|||...|||...
Db 332 ThrSerIleAlaLeuValGlyGlyPheIleLeuValMetGlyPheMetValLeuTyr 351
QY 341 TATAGATTGGAGCGTCATCGCTCGGAGCTCTTCTCTGAATCTTTGCTTATCTGG 400
    |||...|||...|||...|||...|||...|||...|||...|||...|||...
Db 352 TyrSerMetAlaGlyValIleAlaCysLeuAlaLeuValValAsnLeuPheLeuVal 371
QY 401 GCAGCTCTACAGTATTTGGATGCCCACTACCTTGTTCAGGACTCGCTGGGATTTCTT 460
    |||...|||...|||...|||...|||...|||...|||...|||...|||...
Db 372 AlaValMetAlaIlePheGlyAlaThrLeuThrLeuProGlyMetAlaGlyIleValLeu 391
QY 461 GCTATGGGATGGCGGTAGATGCAAAATGTTCTGTATTCGAAAGAACCCGAGAGAAATTT 520
    |||...|||...|||...|||...|||...|||...|||...|||...|||...
Db 392 ThrValGlyIleAlaValAspAlaAsnIleIleAsnGluArgIleArgGluValLeu 411
QY 521 TTATTGTCTCAAGCTCTTAAATCTGTAGAAAAGATATACCAAGGCTTTTGGAGCC 580
    |||...|||...|||...|||...|||...|||...|||...|||...|||...
Db 412 ArgGluAsnGluGlyIleAlaLysAlaIleHisLeuGlyTyrIleAsnAlaSerArgAla 431
QY 581 ATTTTGTATCTCACTTACGATGATGATGATGATGATGATGATGATGATGATGATGAT 640
    |||...|||...|||...|||...|||...|||...|||...|||...|||...
Db 432 IlePheAspSerAsnIleThrSerLeuIleAlaSerValLeuLeuTyrAlaTyrGlyThr 451
QY 641 GGGCTATTAAGGTTGCTTTCACATGATGATTTTGAAGATTTTCTTCAATGTTTACG 700
    |||...|||...|||...|||...|||...|||...|||...|||...|||...
Db 452 GlyAlaIleLysGlyPheAlaLeuThrThrGlyIleGlyIleLeuAlaSerIleIleThr 471
QY 701 GCTCTTTTCATGACTAAATTTTCTTCATGCTGTGGATGATGATGATGATGATGATGAT 760
    |||...|||...|||...|||...|||...|||...|||...|||...|||...
Db 472 AlaIleValGlyThrGlnGlyIleTyrGlnAlaLeuLeuProLysLeuThrGlnThrLys 491

RESULT 3
ID _SECD_HELPJ STANDARD; PRT; 526 AA.
AC Q9ZJ66;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Protein-export membrane protein.
GN _SECD OR JHP1449.
OS Helicobacter pylori J99 (Campylobacter pylori J99).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=85963;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99120557; PubMed=9923682;
RA Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,
RA Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,
RA Tummino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,
RA Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,
RA Trust T.J.;
RT "Genomic sequence comparison of two unrelated isolates of the human
RL gastric pathogen Helicobacter pylori."
RL Nature 397:176-180(1999)
CC -1- FUNCTION: INVOLVED IN PROTEIN EXPORT (BY SIMILARITY).
CC -1- SUBUNIT: PART OF THE PROKARYOTIC PROTEIN TRANSLOCATION APPARATUS
CC WHICH COMPRISE SECA, SECD, SECE, SECF, SECG AND SECY
CC (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).
CC -1- SIMILARITY: BELONGS TO THE SECD/SECF FAMILY. SECD FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
```

```
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements/
CC or send an email to license@ebi.ac.uk).
CC -----
DR EMBL; AB001567; AD07024.1; -.
DR InterPro; IPR001036; ACRFLVIN_res.
DR InterPro; IPR003335; SecD SecF.
DR Pfam; PF02355; SecD SecF.
DR PRINTS; PR00702; ACRFLVAVINRP.
DR TIGRFAMs; TIGR00916; 2A0604s01; 1.
DR TIGRFAMs; TIGR01129; secD; 1.
KW Protein transport; Translocation; Transmembrane; Membrane;
KW Complete proteome.
FT TRANSMEM 8 28 POTENTIAL.
FT TRANSMEM 356 376 POTENTIAL.
FT TRANSMEM 379 399 POTENTIAL.
FT TRANSMEM 453 473 POTENTIAL.
FT TRANSMEM 478 498 POTENTIAL.
SQ SEQUENCE 526 AA; 56796 MW; 5B4C9E905794782A CRC64;
```

```
Alignment Scores:
Pred. No.: 6,79e-27 Length: 526
Score: 429.00 Matches: 97
Percent Similarity: 60.83% Conservative: 49
Best Local Similarity: 40.42% Mismatches: 92
Query Match: 12.95% Indels: 2
DB: 1 Gaps: 2
```

US-09-868-987-1 (1-1864) x SECD_HELPJ (1-526)

```
QY 47 GGCAATATTCT--GCAAAACCGTGAAGTCAGTATGCTGTAGTATTGACGGTTATATG 103
    |||...|||...|||...|||...|||...|||...|||...|||...|||...
Db 274 GlyAspPheSerGlyAlaAsnValGlyLysArgMetAlaIleValLeuAspAsnLysVal 293
QY 104 GTGAGCAGCCCTATTTTAAACGTCCTCATTC--AAAATCATGCCAGTGTCTCAGGAAA 160
    |||...|||...|||...|||...|||...|||...|||...|||...|||...
Db 294 TyrSerAlaProValIleArgGluArgIleGlyGlySerGlyGlnIleSerGlyAsn 313
QY 161 TTTACCCACCGTGAAGTCAGCAAACTGCCTCAGATTTAAATCTGAGCGATCTCTTT 220
    |||...|||...|||...|||...|||...|||...|||...|||...|||...
Db 314 PheSerValAlaGlnAlaSerAspLeuAlaLeuAlaLeuArgSerGlyAlaMetAsnAla 333
QY 221 GTTCCCGAGGTTCTCAGTCAGAGACGATCTCTCTGATCTTTGGGAAAAACAATGTACA 280
    |||...|||...|||...|||...|||...|||...|||...|||...|||...
Db 334 ProIleGlnValLeuGluLysArgIleValGlyProSerLeuGlyLysAspSerIleLys 353
QY 281 CAAGGCATTATCTCAGCATGCTGTGGCTGGCAATGCTTATTTGATGAGCGTATAT 340
    |||...|||...|||...|||...|||...|||...|||...|||...|||...
Db 354 ThrSerIleAlaLeuValGlyGlyPheIleLeuValMetGlyPheMetAlaLeuTyr 373
QY 341 TATAGATTGGAGCGTCATCGCTCGGAGCTCTTCTCTGAATCTTTGCTTATCTGG 400
    |||...|||...|||...|||...|||...|||...|||...|||...|||...
Db 374 TyrSerMetAlaGlyValIleAlaCysMetAlaLeuValValAsnLeuPheLeuVal 393
QY 401 GCAGCTCTACAGTATTTGGATGCCCACTACCTTGTTCAGGACTCGCTGGGATTTCTT 460
    |||...|||...|||...|||...|||...|||...|||...|||...|||...
Db 394 AlaValMetAlaIlePheGlyAlaThrLeuThrLeuProGlyMetAlaGlyIleValLeu 413
QY 461 GCTATGGGATGGCGGTAGATGCAAAATGTTCTGTATTCGAAAGAACCCGAGAGAAATTT 520
    |||...|||...|||...|||...|||...|||...|||...|||...|||...
Db 414 ThrValGlyIleAlaValAspAlaAsnIleIleAsnGluArgIleArgGluValLeu 433
QY 521 TTATTGTCTCAAGCTCTTAAATCTGTAGAAAAGATATACCAAGGCTTTTGGAGCC 580
    |||...|||...|||...|||...|||...|||...|||...|||...|||...
Db 434 ArgGluGlyGluGlyValValLysAlaIleHisLeuGlyTyrIleAsnAlaSerArgAla 453
QY 581 ATTTTGTATCTCACTTACGATGATGATGATGATGATGATGATGATGATGATGATGAT 640
    |||...|||...|||...|||...|||...|||...|||...|||...|||...
Db 454 IlePheAspSerAsnIleThrSerLeuIleAlaSerValLeuLeuTyrAlaTyrGlyThr 473
```


QY 29 ATCAGCGGCACT---GCTAATGAGCAATATCTGCAACCGT---GGATGGCGTATGGCT 82
 Db 305 ValSerGlyAlaArgAlaPheGlyAspTryThrAlaGlyHisIleGlyGlnProPheAla 324
 QY 83 GTATGATGATGAGCTGTATATGCTGACAGACCTTATTTAAACGTCCTCAAAAAATCAT 142
 Db 325 IleValLeuAspGlyLysValIleSerAlaProThrIleGlnAlaHisIleAlaGly 344
 QY 143 GCCAGT---GTCTCAGGGAATTTACCAACCGTGAAGTAGCAAACTCGCTCAGATT 199
 Db 345 SerGlyIleIleThrGlyArgPheSerIleGlnGlnAlaThrAspLeuAlaLeuLeu 364
 QY 200 AAATCTGAGAGGATGCTTCTTGTCCGAGGTCTCAGTGAAGAGAGATCTTCTGAT 259
 Db 365 ArgAlaGlyAlaLeuProAlaGlyMetThrPheLeuGlnGlyThrIleGlyProGlu 384
 QY 260 CTTGGAGAAAAACATGATACACAGGCAATATCTCAGCATCTGCTGGCTGGCAATGCT 319
 Db 385 LeuGlyAlaAspSerValLysAlaGlyMetValAlaSerValIleGlyPheValAlaVal 404
 QY 320 ATGTGTTGATGAGCGTATATTATATGATTTGAGGCGTATCGCTTGGAGCTGTTCTT 379
 Db 405 ValAlaTryMetIleAlaSerTrycLysPheGlyPhePheSerSerValAlaLeuPhe 424
 QY 380 CTGAATCTTTGCTTATCTT3GGCAGCTCTACAGTATTGGATGGCCACTACCTTGCA 439
 Db 425 IleAsnIleAlaPheIlePheAlaValMetGlyAlaIleGlyGlyThMetThrLeuPro 444
 QY 440 GGACGCGCGGATGTTGCTTGTCTGCTATGAGGATGCGCTAGATGCAATGTTGTTATTC 499
 Db 445 GlyIleAlaGlyIleValLeuThrIleGlyThrSerValAspAlaAsnValLeuIleYr 464
 QY 500 GAAAGATCCAGAGAGAAATTTTATTTGCTCAAGTCTTAAATAATGTAAGAAAAAGA 559
 Db 465 GlnArgMetArgGlnGluIleArgSerGlySerProValArgAlaIleGlnLeuGly 484
 QY 560 TATACCAAGCTTTTGAGCCATTTTATTTGATTTGATCTGACTACAGTATTGGCCCTACA 619
 Db 485 PheAspLysAlaMetSerAlaIleIleAspAlaAsnValThrSerPheLeuSerSerAla 504
 QY 620 CTTGTTTCTTCTAGATACAGAGGCTTATTAAGGGTTTGCTTGAATGATTTAGGA 679
 Db 505 IleLeuPheValLeuGlyAlaGlyProValArgGlyPheAlaValThrThrMetIleGly 524
 QY 680 ATTTTCTCTCATGTTTACGCTCTTTTCACTGACTAAATTTTCTTCACTGCTGG 736
 Db 525 IleAlaIleSerIlePheThrAlaIleThrPheValValArgLeuMetIleValIleThr 543
 RESULT 7
 SECD_HAEIN
 ID SECD_HAEIN STANDARD; PRT; 616 AA.
 AC P44591;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Protein-export membrane protein secd.
 GN SECD OR H10240.
 OS Haemophilus influenzae.
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
 OC Haemophilus.
 OX NCBI_TaxID=727;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Rd / KM20 / ATCC 51907;
 RX MEDLINE=95350630; PubMed=7542800;
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
 RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
 RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
 RA Scott J.D., Shireley R., Liu L.-I., Glodek A., Kelley J.M.,
 RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom B., Cotton M.D.,
 RA Uettermann T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
 RA Fine L.D., Fitchman J.L., Fuhmann J.L., Geoghagen N.S.M.,
 RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,

RA Venter J.C.;
 RT "whole-genome random sequencing and assembly of Haemophilus influenzae
 RT R1";
 RL Science 269:496-512(1995).
 CC -1- FUNCTION: INVOLVED IN PROTEIN EXPORT (BY SIMILARITY).
 CC -1- SUBUNIT: PART OF THE PROKARYOTIC PROTEIN TRANSLLOCATION APPARATUS
 CC WHICH COMPRISE SECA, SECB, SECD, SECE, SECF, SECG AND SECH
 CC (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
 CC (By similarity).
 CC -1- SIMILARITY: BELONGS TO THE SECD/SECF FAMILY. SECD FAMILY.
 CC
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: U32710; AAC21908.1; -.
 CC TIGR: H10240; -.
 DR InterPro: IPR003335; Secd_SecF.
 DR Pfam: PF02355; Secd_SecF_1.
 DR TIGRFAMs: TIGR00916; 2A0604501; 1.
 DR TIGRFAMs: TIGR01129; secd; 1.
 KW Protein transport; Translocation; Transmembrane; Inner membrane;
 KW Complete proteome.
 FT TRANSMEM 11 31 POTENTIAL.
 FT TRANSMEM 453 473 POTENTIAL.
 FT TRANSMEM 475 495 POTENTIAL.
 FT TRANSMEM 497 517 POTENTIAL.
 FT TRANSMEM 562 582 POTENTIAL.
 FT TRANSMEM 585 605 POTENTIAL.
 SQ SEQUENCE 616 AA; 66986 MW; 7DE1E4B08505F8F CRC64;
 Alignment Scores:
 Pred. No.: 3.1e-25 Length: 616
 Score: 408.50 Matches: 87
 Percent Similarity: 65.85% Conservative: 48
 Best Local Similarity: 42.44% Mismatches: 59
 Query Match: 12.33% Indels: 11
 DB: 1 Gaps: 2
 US-09-868-987-1 (1-1864) x SECD_HAEIN (1-616)
 QY 137 AATCATCCAGTGTCTCAGGAAATTT-----ACC 166
 Db 395 AsnValAlaThrIleGlnGlyArgPheGlySerAsnPheGlnIleThrGlyValAspSer 414
 QY 167 CACCGTAAGTGAACAACTGCGCTCAGATTAAATCTGGAGGATGCTTTGTGCC 226
 Db 415 IleAlaGlnAlaHisAsnLeuSerThrLeuLeuYserGlyAlaLeuIleAlaProIle 434
 QY 227 GAGGTTCTCAGTGAAGAGAGATCTCTTGATCTTGGGAAAAACATGATACAGGC 286
 Db 435 GlnIleValGlnGluArgThrIleGlyProSerLeuGlyAlaGlnAsnValGlnGlnGly 454
 QY 287 ATATCTCAGATGCTGTGCTTGGCAATGCTTATTTGATGAGCGATATTATATAGA 346
 Db 455 IleAsnAlaSerLeuThrPgsLysLeuValAlaValIleAlaPheMetLeuPheTryTyls 474
 QY 347 TTGGAGGCGTATCGCTTGGGAGCTGTTCTTGAATCTTTGCTTATCTGGGAGCT 406
 Db 475 MetPheGlyValIleAlaSerPheAlaLeuValIleAsnIleValLeuLeuValGlyLeu 494
 QY 407 CTACAGTATTG---GATGGCCACTCATCTTGACAGACTCGCTGGATGTTCTTGCT 463
 Db 495 MetSerIleLeuProGlyAlaThrLeuSerMetProGlyIleAlaGlyIleValLeuThr 514
 QY 464 ATGGGAGTGGCGGAGATGCAAAATGTTCTTGATTCGAAAGAAATCCGAGAGGAATTTTA 523
 Db 515 LeuGlyMetSerValAspAlaAsnValLeuIlePheGlnArgIleLysGlnGlnIleArg 534

QY 524 TTGTCCTCAAGTCTTAAATACTGTAGAAAAGGATATACCAAGGCTTTTGGAGCAATT 583
 Db 535 AsnGlyArgSerIleGlnAlaIleAsnGluGlyTyrAsnGlyAlaPheThrSerIle 554
 QY 584 TTTGATCTTAACCTGACTACAGTATTGGCTCAGCACTTCTTTCTTCCTAGATACAGGG 643
 Db 555 PheAspAlaAsnLeuThrIleLeuThrAlaIleLeuThrAlaIleLeuThrAlaIleGlyThrGly 574
 QY 644 CTTATTAAAGGTTGCTTTGACATTTAGGAAATTTCTTCAATGTTTACGGCT 703
 Db 575 ProIleGlnGlyPheAlaIleThrLeuSerLeuGlyValAlaIleSerMetPheThrAla 594
 QY 704 CTTTTCATGACTAAA 718
 Db 595 IleThrGlyThrArg 599

RESULT 8
 ID SECD_BRUAB STANDARD; PRT; 518 AA.
 AC Q9ZG86;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE Protein-export membrane protein secD (Fragment).
 GN SECD.
 OS Brucella abortus.
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Brucellaceae; Brucella.
 OX NCBI_TaxID=235;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=2308;
 RX MEDLINE=99043892; PubMed=9826342;
 RA Vemulapalli R., Duncan A.J., Boyle S.M., Sritanganathan N., Toth T.E.,
 RA Schurig G.G.;
 RT "Cloning and sequencing of yajC and secD homologs of Brucella abortus
 RT and demonstration of immune responses to yajC in mice vaccinated with
 RT B. abortus RB51.";
 RL Infect. Immun. 66:5684-5691(1998).
 CC -1- FUNCTION: INVOLVED IN PROTEIN EXPORT (BY SIMILARITY).
 CC -1- SUBUNIT: PART OF THE PROKARYOTIC PROTEIN TRANSLOCATION APPARATUS
 CC WHICH COMPRISE SECA, SECB, SECD, SECE, SECF, SECG AND SECH
 CC (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).
 CC -1- SIMILARITY: BELONGS TO THE SECD/SECF FAMILY. SECD FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AF085217; AAC83691.1; -
 CC InterPro; IPR001036; Acrlflvln_res.
 CC InterPro; IPR003335; SecD SecF.
 CC Pfam; PF02335; SecD SecF.1.
 CC PRINTS; PR00702; ACRIFLAVINRP.
 CC TIGRFAMS; TIGR00916; 2A0604801; 1.
 CC TIGRFAMS; TIGR01129; secD; 1.
 CC Protein transport; Translocation; Transmembrane; Membrane.
 CC TRANSMEM 9 29 POTENTIAL.
 CC TRANSMEM 277 297 POTENTIAL.
 CC TRANSMEM 301 321 POTENTIAL.
 CC TRANSMEM 392 412 POTENTIAL.
 CC TRANSMEM 460 480 POTENTIAL.
 CC NON_TER 518 518
 CC SEQUENCE 518 AA; 55685 MW; 35250B66429A7B3B CRC64;

Alignment Scores:
 Pred. No.: 3.66e-25 Length: 518

Score: 407.50 Matches: 96
 Percent Similarity: 56.58% Conservative: 76
 Best Local Similarity: 31.58% Mismatches: 107
 Query Match: 12.30% Indels: 25
 DB: 1 Gaps: 7

US-09-868-987-1 (1-1864) x SECD_BRUAB (1-518)

QY 38 ACTGCTAAATGGACAATATTCTGCACAAACCGTGGATGCGGTATGCTGTAGTATGACGGT 97
 Db 202 ThrAlaGlnGlyAsnGluAsnSer-----PheAlaIleValAlaAspAsn 216
 QY 98 TATATGCTCAGCAGCCCTATTATTAACGTCCTCCATTG---AAAAATCATGCCAGTGCTCA 154
 Db 217 GlnValValSerAlaAlaProThrValSerGlyProLeuAspThrSerGluLeuGlnIleGlu 236
 QY 155 GGGAAATTTTACCACCGTGAAGTGAGCAAACTCGCTCAGATTTAAATCTGAGCGCATG 214
 Db 237 GlyAlaPheAspLeuGlnAlaAlaAsnAsnMetAlaValValLeuArgSerGlyAlaLeu 256
 QY 215 TCTTTTGTCCCGAGGTTCTCAGTGAAGAGACGATCTTCTGATCTTGGGAAACAA 274
 Db 257 ProGlnAlaValThrValLeuGluArgThrIleAlaSerAlaLeuGlyGluAspTyr 276
 QY 275 TGTACACAGGCATTATCTCAGCATCTGTGGCTTGGCAATGCTTATGTTTGTATGAGC 334
 Db 277 AlaSerAlaAlaValLeuAlaAlaLeuLeuAlaAlaValValGlyLeuPheMetVal 296
 QY 335 GTATATTATAGATTGAGGCGTCATCGTTCGGGAGCTGTTCTCTGAATCTTTGCTT 394
 Db 297 LeuSerTyrGlyIleLeuGlyValIleAlaLeuValAlaValValAlaIleValIle 316
 QY 395 ATCTGGGAGCTCTCACAGTATTGGATGGCCACTCACCTTGTGACGAGCTCGCTGGGATT 454
 Db 317 LeuThrAlaValLeuSerLeuIleGlyAlaSerIleSerLeuAlaSerIleAlaGlyLeu 336
 QY 455 GTTCTTGCTATGGGATGGCGTAGATGCAATGTTCTTGTATTGCAAGAAATCCGAGAG 514
 Db 337 ValLeuThrIleGlyLeuAlaValAlaPheAlaHiaIleLeuIleTyrGluArgValArgGlu 356
 QY 515 GAATTTTATTGCTCAAGTCTTAAATACTGTAGAAAAGGATATACCAAGGCTTTT 574
 Db 357 AspArgArgGlyTyrSerValValGlnAlaMetGluSerGlyPheTyrArgAlaLeu 376
 QY 575 GGAGCCATTTTGTATCTTAACCTGACTACAGTATTGGCTCAGCACTTCTTTTCTCTTA 634
 Db 377 SerThrIleValAspAlaAsnLeuThrThrLeuIleAlaAlaLeuValLeuPheLeuLeu 396
 QY 635 GATACAGGCGCTATTAAAGGTTGCTTGCATTTAGGAAATTTCTCTCAATG 694
 Db 397 GlySerGlyThrValHisGlyPheAlaLeuThrValAlaIleGlyIleGlyThrThrLeu 416
 QY 695 TTTACGGCTCTTTTCATGACTAAATTTTCTTCATGCTGTGGATGAAT-----NAGACC 748
 Db 417 PheThrThrLeuThrPheThrArgLeuLeuIleAlaGlnTrpValArgThrAlaLysPro 436
 QY 749 CAACATACACAGTTGCATATATGATGAATAAGTTCGTGGGATAAAGCAT---GATTTCTTG 805
 Db 437 LysGluValProLysArgArgLeuLysLeuValProThrValThrHisIleProPheMet 456
 QY 806 AGAGGATGCAAAAAAATTTGGGCTGTTCTGGAAGTGTTTCTTTTAGGTTCG-----859
 Db 457 Arg-----LeuGlnPheValThrLeuGlyIleSerValLeuAlaCysAlaIle 472
 QY 860 -----GTTGCTCTCGGGTTTGGAGCTCGGAATTCGGTTTGGGAATGGAT 904
 Db 473 ValValAlaLeuPheValAlaIleGlyPheAsn-----TyrGlyIleAsp 487
 QY 905 TTTAAAGGAGGG 916
 Db 488 PheArgGlyGly 491

RESULT 9

```

SECD_RICPR
ID SECD_RICPR STANDARD: PRT; 518 AA.
AC Q92CW8;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Protein-export membrane protein secd.
OS SECD OR RP586.
OS Rickettsia prowazekii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiaseae; Rickettsia.
OX NCBI_TaxID=782;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Madrid E;
RX MEDLINE=99039499; PubMed=9823893;
RA Andersson S.G.E., Zomorodipour A., Andersson J.O.,
RA Sierheltz-Ponten T., Almark U.C.M., Podowski R.M., Naeslund A.K.,
RA Eriksson A.-S., Winkler H.H., Kurland C.G.;
RT "the genome sequence of Rickettsia prowazekii and the origin of
RT mitochondria."
RL Nature 396:133-140(1998).
CC -1- FUNCTION: INVOLVED IN PROTEIN EXPORT.
CC -1- SUBUNIT: PART OF THE PROKARYOTIC PROTEIN TRANSLOCATION APPARATUS
CC WHICH COMPRISE SECA, SECB, SECD, SECE, SECF, SECG AND SECH
CC (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (by similarity).
CC -1- SIMILARITY: BELONGS TO THE SECD/SECF FAMILY. SECD FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AJ235272; CAA15031.1; -.
DR InterPro; IPR003335; SecD_SecF.
DR Pfam; PF02355; SecD_SecF.1.
DR TIGRFAMs; TIGR00916; 2A0604s01.1.
DR TIGRFAMs; TIGR01129; secd.1.
KW Protein transport; Translocation; Transmembrane; Membrane;
KW Complete proteome.
FT TRANSMEM 9
FT TRANSMEM 29
FT TRANSMEM 356
FT TRANSMEM 376
FT TRANSMEM 377
FT TRANSMEM 397
FT TRANSMEM 406
FT TRANSMEM 426
FT TRANSMEM 464
FT TRANSMEM 484
FT TRANSMEM 486
FT TRANSMEM 506
FT TRANSMEM 518
SQ SEQUENCE 518 AA; 56712 MW; 86DFC2A5B5DE2AA CRC64;

Alignment Scores:
Pred. No.: 1,956-24 Length: 518
Score: 398.50 Matches: 86
Percent Similarity: 58.8% Conservativeness: 54
Best Local Similarity: 36.1% Mismatches: 97
Query Match: 12.0% Indels: 1
Gaps: 1

US-09-868-987-1 (1-1864) x SECD_RICPR (1-518)
QY 62 AACCGTGATGCGGTAGTGTAGTGAAGTATATGTCAGACCCCTATTTTA 121
DB 281 AenThnrglyLysArgLeuAlaValleuAspAsnLysLeuLeuSerAlaProIleIle 300
QY 122 AAC---GTCCTGATGAAAATATATGCGCATGCTCAGGAAATTTTACCAACCGTGAAG 178
DB 301 AsnglyAlaIleIleGlyCylAsnGlyIleIleThnrglyAsnPhethrIleGluSerAla 320
QY 179 AGCAACTGCGCTGATTAATAATCTGGAGGAGTCTTTTGTCCGAGGTTCTCACT 238
DB 321 AsnGluLeuAlaLeuLeuLeuAlaValGlySerLeuProThrProLeuLysIleIleGlu 340

```

```

QY 229 GAGAGACGATCTCTTGATCTTGGAAAAAATGTATACAGGCAATTATCTCACA 298
DB 341 GlnTrsSerIleGlyProAsnLeuGlyAlaAspSerIleGluSerGlyLysAlaIle 360
QY 229 TGCTGTGCTTGGCAATCTTATTTTGTATGATGAGCGATATATPAGATTGGAGCGTC 358
DB 361 LeuIleGlyPheValAlaValCysIlePheMetIleLeuSerTyrGlyValIleGlyLeu 380
QY 359 ATCGCTTGGAGAGCTGTTCTTCTGAATCTTTTGTCTTATCTGGGAGCTTACAGATTGG 418
DB 381 PheAlaAsnIleAlaLeuIleLeuAlaLeuLeuAlaLeuLeuSerLeuPhe 400
QY 419 GATGGCCACTCACCCTTGTATGACGACTGCTGGAGATTGTTCTGTATGAGGATGCCGTA 478
DB 401 GlnAlaThrLeuThrLeuProGlyIleAlaGlyIleIleLeuThrIleGlyMetAlaVal 420
QY 479 GATGCAATGTTCTTGTATGCAAGAAATCCGAGAGGAAATTTTATGTCACAAAGTCTT 538
DB 421 AspAlaAsnValIleIleTyrGlnArgIleLysGlnGluLeuHisLysGlyValSerAsn 440
QY 539 AAAAAATCTGTAGAAAAAGATATACCAAGCCTTTTGGAGCCATTTTGTATCTTACTTG 598
DB 441 LeuTyrAlaIleArgThnrglyPheGluSerAlaPheAlaThrIleIleAspSerAsnIle 460
QY 599 ACTACAGATTGGCGCTCAGCACTTCTTTCTTCTAGATATACAGGCGCTATTAAGGGTTT 658
DB 461 ThrThrIleValAlaPheAlaLeuThrIlePheGlyValGlyAlaIleLysGlyPhe 480
QY 659 GCTTGTGATGATGATTTTGTAGAAATTTTCTCAATGTTTACGGCTTTTCATGACTTAA 718
DB 481 AlaValAlaLeuThrIleGlyIleIleSerSerMetPheSerAlaIleIleThrIlys 500
QY 719 TTTTCTTCATGCTGTGATGATTAAGCCACATACACAGTTTGATATGATG 772
DB 501 LeuLeuIleAspValTrrpValLysTyrPheLysProLysLysLeuGlyLeuLeu 518

RESULT 10
ID SECF_ECOLI STANDARD: PRT; 323 AA.
AC P19674; P71113;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Protein-export membrane protein secf.
GN SECF OR B0409 OR Z0508 OR ECS0460.
OS Escherichia coli, and
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562, 83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=9106014; PubMed=2170107;
RA Gardel C., Johnson K., Jacq A., Beckwith J.;
RT "The secd locus of E.coli codes for two membrane proteins required
RT for protein export."
RL EMBO J. 9:3209-3216(1990).
RN [2]
RP ERRATUM.
RX MEDLINE=9106534; PubMed=2249673;
RA Gardel C., Johnson K., Jacq A., Beckwith J.;
RL EMBO J. 9:4205-4206(1990).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blatter F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-VIDES J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.U.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."

```


DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE Protein-export membrane protein secf.
 GN SECF.
 OS Rhodobacter capsulatus (Rhodospirillum rubrum).
 OC Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;
 OC Rhodobacter.
 OX NCBI_TaxId=1061;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SB1003 / St Louis;
 RX MEDLINE=97318920; PubMed=9175857;
 RA Goldman B.S., Beckman D.L., Ball A., Monika E.M., Gabbert K.K.,
 RA Kranz R.G.;
 RT "Molecular and immunological analysis of an ABC transporter complex
 RT required for cytochrome c biogenesis."
 RL J. Mol. Biol. 268:724-738(1997).
 CC -1- FUNCTION: INVOLVED IN PROTEIN EXPORT.
 CC -1- SUBUNIT: PART OF THE PROKARYOTIC PROTEIN TRANSLOCATION APPARATUS
 CC WHICH COMPRISE SECF, SECB, SECD, SECE, SECF, SECG AND SECY
 CC (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).
 CC -1- SIMILARITY: BELONGS TO THE SECD/SECF FAMILY. SECF FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/ebis-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 DR EMBL: U69979: AAB62802.1; -;
 DR InterPro: IPR003335; SecD_SecF.
 DR Pfam: PF02355; SecD_SecF; 1.
 DR TIGRPFAMs: TIGR00916; 2A0604801; 1.
 DR TIGRPFAMs: TIGR00966; 3A0501807; 1.
 KM Protein transport; Translocation; Transmembrane; Membrane.
 FT TRANSMEM 27 47 POTENTIAL.
 FT TRANSMEM 152 172 POTENTIAL.
 FT TRANSMEM 180 200 POTENTIAL.
 FT TRANSMEM 271 291 POTENTIAL.
 SQ SEQUENCE 333 AA; 36777 MW; 4DF5FAF12A0FE073 CRC64;
 Alignment Scores:
 Pred. No.: 1.95e-21 Length: 333
 Score: 361.00 Matches: 100
 Percent Similarity: 43.94% Conservative: 56
 Best Local Similarity: 28.17% Mismatches: 111
 Query Match: 10.90% Indels: 88
 DB: 1 Gaps: 8
 US-09-868-987-1 (1-1864) x SECF_RHOCA (1-333)
 QY 797 GATTCTTGAAGATGCAAAAACCTTGGCGCTTTCTGGA----- 838
 DB 15 Asphephehargtrp-----GlntrpAlahrhphgelyAlaAlaIleValMetMet 31
 QY 839 -----AGTGTCTTTCTTTAGGTGGCTGCTCGGGTTTGAGCGCTGGAATTCGGT 892
 DB 32 lIleAlaSerValIleLeuProLeuValIleGlyLeuAsnPh----- 45
 QY 893 TTGGGAATGATTTTAAAGGAGG----- 916
 DB 46 ---GlylleAspPhehlyGlyGlyThrThrIleArgThrGluSerThrThrAlaIleAsp 64
 QY 917 -----TATGCTTATACCTTAAATCAAAAGAGCATGCGATCAGATGTCGCA 967
 DB 65 ValGlyValIlyrArgAlaAlaLeuGluProLeuGluLeuGlyAspValIleIleSerGlu 84
 QY 968 ATGCGT-----GGCAAGTTTGTCATTAACATACAG 997
 DB 85 ValArgAspProSerPheArgGluAspGluAspGlnHisValAlaMetIleArgIleGlnMetGln 104

QY 998 GAAGCTGCTCTTCTCTAGAGACTTCCGATTAACAATTGGATCTTCAGAAAGATC 1057
 DB 105 GlnAspGlyGlnGlyAlaGlnGlyGlnGlyAlaGlnGlyGlnGlyVal 124
 QY 1058 AAAATCTATTTAGT-----GATMAAGCTTTAAGCTATATAGAGATACGACCTCT 1111
 DB 125 GluThrAlaLeuThrAlaValAlaAspProAlaLeuLysIleThrSer----- 139
 QY 1112 CTCCTAAATTAACGATCATGACCTGGCGTTATTGGGGATTGTTGTTCAGAAACAGGCT 1171
 DB 139 ----- 139
 QY 1172 AGATTCTCAGGAAACTTAACGAAAGCAAAATTTGTCGCAAGTAAGCAGCA 1231
 DB 140 -----PheGluSerValGlyProLys 146
 QY 1232 CTATCGAAGAAATGCGTTATCAGCGACCATCGGCTTTTGAAGATCATCATC 1291
 DB 147 ValSerGlyGluLeuValIlyrThrAlaValAlaIlyrSerLeuLeuAlaIlyrValIle 166
 QY 1292 TTGCTTAATGATGATTTGCGCTTTGAATGCAATATGCTTTCAGTCCGATGCGCTTTA 1351
 DB 167 MetPheThrIleThrValAlaArgPheGlnIlyrGlnPheAlaLeuGlyAlaValAlaLeu 186
 QY 1352 ATTCATGACCTTTTGGCTACCTGTCGACGCTTGTATTATACCATTTCTTTGAAGAA 1411
 DB 187 ValHisAspValLeuThr-----ValGlyLeuPheAlaValLeuGln 201
 QY 1412 ATTCATGATTTTGAAGCATGCTGCTTATATGATGATGATGATGATGATGATGATGAT 1471
 DB 202 LeuLysPheAspLeuThrThrValAlaAlaLeuLeuThrIleThrGlyIlyrSerIleAsn 221
 QY 1472 AATACCTTATGATCTTTTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1531
 DB 222 AspThrValValAlaValPheAspArgLeuArgGluAsnLeuLysIlyrIlyrSerMetPro 241
 QY 1532 ATGCACTGTTTATGATATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1591
 DB 242 LeuArgAspAlaMetAsnLeuSerValAsnGlnIlyrThrLeuSerArgThrValMetThrGly 261
 QY 1592 GCTACACCTCTATCAGTTTGTGTTAATGCTTTTATGAGCGGCTCTCTCTTAAAT 1651
 DB 262 MetThrThrLeuLeuAlaLeuValPrometLeuIleIlyrGlyGlyAspValIleArgGly 281
 QY 1652 TTTCATTTATTTATGACCATGAGGATTTCTTAGAATCTTATGCTCTTTATTTATGCA 1711
 DB 282 PheValPheAlaMetValIlyrGlyValIlyrPheThrGlyThrIlyrSerSerValIlyrValAla 301
 QY 1712 CCACCTCTGTTGTTGTTTATGCTGCTTAAAGAAATGCGTCGCAAA 1756
 DB 302 LysAsnIleValAlaLeuPheIleGlyLeuAspArgAsnIlyrGlyLys 316
 RESULT 12
 SECF_RHOSH STANDARD; PRT; 324 AA.
 ID SECF_RHOSH
 AC 033568;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE Protein-export membrane protein secf.
 GN SECF.
 OS Rhodobacter sphaeroides (Rhodospirillum rubrum).
 OC Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;
 OC Rhodobacter.
 OX NCBI_TaxId=1063;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 17023 / 2.4.1 / NCIB 8253 / DSM 158;
 RA Patterson C.S., Donohue T.J.;
 RT "Identification and characterization of putative cytochrome c
 RT maturation genes (cmhABCDG) from Rhodobacter sphaeroides";
 RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: INVOLVED IN PROTEIN EXPORT.

CC -1- SUBUNIT: PART OF THE PROKARYOTIC PROTEIN TRANSLOCATION APPARATUS
 CC WHICH COMPRISE SECA, SECB, SECD, SECE, SECF, SECG AND SECH
 CC (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).
 CC -1- SIMILARITY: BELONGS TO THE SECD/SECF FAMILY. SECF FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; U83136; AAB61904.1; .
 DR InterPro; IPR003335; SecD_SecE.
 DR Pfam; PF02335; SecD_SecF; 1.
 DR TIGRFAMs; TIGR00916; 2A0604s01; 1.
 DR TIGRFAMs; TIGR00966; 3A0501s07; 1.
 KW Protein transport; Translocation; Transmembrane; Membrane.
 FT TRANSMEM 16 36 POTENTIAL.
 FT TRANSMEM 146 166 POTENTIAL.
 FT TRANSMEM 174 194 POTENTIAL.
 FT TRANSMEM 266 286 POTENTIAL.
 SQ SEQUENCE 324 AA; 35088 MW; F4CD14CSB814370E CRC64;
 Alignment Scores:
 Pred. No.: 2,34e-21 Length: 324
 Score: 360.00 Matches: 100
 Percent Similarity: 49.06% Conservative: 56
 Best Local Similarity: 31.45% Mismatches: 100
 Query Match: 10.87% Indels: 62
 DB: 1 Gaps: 10
 US-09-868-987-1 (1-1864) x SECF_RHOSH (1-324)
 Qy 839 AGTGTTTCTTTAGTGGCTGCTCTCGGTTTGGAGCTGAATTCGGT----- 892
 Db 27 SerValPheLeuMet-----AlaSerLeuValAlaTrpLeuThrLeuGlyLeu 43
 Qy 893 ---TTGGCAATGAGTTTAAAGGGGATGCCCTTTACCTTTAATCCAAAGAGCATGCC 949
 Db 44 AsnPheGlyIleAspPheArgGly-----ThrThrIleArgThrGluSerThrGln 61
 Qy 950 ATCAGCGATGTTGCTCAATCGTGGCAAGTTGTGCATAAAGTACAGGAAGCTGCTTT 1009
 Db 62 AlaValAspValAlaAlaTyArg-----AlaAlaLeu 72
 Qy 1010 TCTCTAGAGAC-----TTCCGTATT 1030
 Db 73 GluGlyGlnAspLeuGlyAspIleSerIleThrGluValPheAspProGlyPheArgAla 92
 Qy 1031 CAACATTTGGATCTTCAGAAAGATCAAAATCTATTTTAGTGATAAAGCTTTAGCTAT 1090
 Db 93 AspGlnHisValAla-----MetValArgIleGlyAlaGlnAspAlaThrGlnSerIle 110
 Qy 1091 ACTAAG-----CAGATCAGCGCTCTCTCTCTAAATAATTAACGATCATGAGCTGG 1138
 Db 111 ThrProGluGlnIleGlyGlnValGluGluAlaLeuLysThrValAsp----- 126
 Qy 1139 CGTTATTCTGGATTGTTGTTCAGAAACAGGCGCTAGATTCTCTACGGAACCTCTAAACGA 1198
 Db 127 -----ProSerIleThrPheProSerValGlu--- 135
 Qy 1199 AACGCAAAATTTTGGTCAAGGTAGCAGCAAACTATCGAAGAAATCGGTTATCAGCGC 1258
 Db 136 -----SerValGlyProLysValSerGlyGluLeuIleArgSerAla 149
 Qy 1259 ACCATCGGGCTTTTAGGCAATCATCTTGGCTCTATGTGAGTTTGGCTTTCGAA 1318
 Db 150 IleLeuAlaValAlaAlaCysAlaGlyIleAlaValTyIleTrpLeuArgGlu 169
 Qy 1319 TGGCAATATGCTTTTCAGTGCCTATGCGCTTTAATTATCATGACCTTTTGGCTACCTGTGCA 1378

Db 170 TrpGlnPheAlaLeuGlySerValAlaAlaLeuIleHisaspValLeuValThrIleGly 189
 Qy 1379 GTCTTGTATTATAGCACATTTCTTTTGAAGAAAATTCAAATAGATTTTCAAGCCCATGGT 1438
 Db 190 Val-----PheAlaLeuPheGlnIleLysPheAspLeuThrThrValAla 204
 Qy 1439 GCTTTAATGACTGATGTTGGGTATTCATTAACAATACTTTTGATCATTTTGTGCTGATT 1498
 Db 205 AlaLeuLeuThrValLeuGlyTySerIleAsnAspThrValValValPheAspArgLeu 224
 Qy 1499 CGTCAAGATCGCCAAAGCAACCTGTTTACCCTATGCATGTTTGTAGTTAATGATGCCCTT 1558
 Db 225 ArgGluAsnLeuValLysTyLysThrMetProLeuArgAspValMetAsnLeuSerVal 244
 Qy 1559 CAAAGACGTTTCAGCCCGCAGTAAATGACACACAGCTACAACTCTATCAGTTTGTTAATG 1618
 Db 245 AsnGluThrLeuSerArgThrIleMetThrLeuMetThrThrLeuIleAlaLeuValSer 264
 Qy 1619 CTTTGTGTTTATAGCGGCTCTCTCTCTTTAAATTTTGCATTTATTATGACCATAGGATT 1678
 Db 265 LeuLeuValPheGlyGlyAspValIleArgGlyPheValPheAlaIleThrPheGlyVal 284
 Qy 1679 CTTCTAGAACTTTATCGTCTCTCTTTATATTCACCACTCTGTTGTTGTTTATG 1732
 Db 285 ValIleGlyThrTySerSerValTyMetAlaLysAsnIleValLeuTyLeu 302
 RESULT 13
 SECD_ECOLI STANDARD; PRT; 615 AA.
 ID AC P19673; P77531; P72348;
 DT 01-FEB-1991 (Rel. 17, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Protein-export membrane protein secd.
 GN SECD OR B0408 OR Z0507 OR ECS0459.
 OS Escherichia coli, and
 OS Escherichia coli O157:H7.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 OX NCBI_TaxID=562, 83334;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=9106014; PubMed=2170107;
 RA Gardel C., Johnson K., Jacq A., Beckwith J.;
 RT "The secd locus of E.coli codes for two membrane proteins required
 RL for protein export.";
 RL EMBO J. 9:3209-3216(1990).
 RN [2]
 RP ERRATUM.
 RX MEDLINE=91065354; PubMed=2245673;
 RA Gardel C., Johnson K., Jacq A., Beckwith J.;
 RL EMBO J. 9:4205-4206(1990).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glaesner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12.";
 RL Science 277:1453-1474(1997).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glaesner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12.";
 RL Science 277:1453-1474(1997).
 RN [5]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / EDL933 / ATCC 700927;

RA MEDLINE=21074951; PubMed=1120651;
 RA Perna N.T., Plunkett G., III, Burland V., Mau B., Glasner J.D.,
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
 RA Postel G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
 RA Postel G.J., Davis N.W., Lim A., Dimalanta E.T., Potamoudis K.,
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
 RA Welch R.A., Blattner F.R.;
 RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7,"
 RL Nature 409:529-533 (2001).
 RN [6]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / RIMD 0509952;
 RX MEDLINE=21156231; PubMed=11258796;
 RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Tanaka M., Tobe T.,
 RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Yaeunaga T.,
 RA Iida T., Takam H., Honda T., Sasaki K., Ogasawara N., Yaeunaga T.,
 RA Kihara S., Shiba T., Hattori M., Shinagawa H.;
 RA "Complete genome sequence of enterohaemorrhagic Escherichia coli
 RT O157:H7 and genomic comparison with a laboratory strain K-12.";
 RL DNA Res. 8:11-22(2001).
 CC [7]
 CC SEQUENCE OF 1-76 FROM N.A.
 RX MEDLINE=94131960; PubMed=7507921;
 RA Pogliano K.J., Beckwith J.;
 RT "Genetic and molecular characterization of the Escherichia coli *secD*
 RT operon and its products.";
 RL J. Bacteriol. 176:804-814(1994).
 CC -I- FUNCTION: INVOLVED IN PROTEIN EXPORT.
 CC -I- SUBUNIT: PART OF THE PROKARYOTIC PROTEIN TRANSLLOCATION APPARATUS
 CC WHICH COMRISES SECA, SECB, SECD, SECE, SECF, SECG AND SECY.
 CC -I- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane.
 CC -I- SIMILARITY: BELONGS TO THE SECD/SECF FAMILY. SECD FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-slb.ch/announce/>
 CC or send an email to license@isb-slb.ch).
 CC
 CC -----
 DR EMBL, X56175; CAA39634.1; -
 DR EMBL, AE00147; AAC73511.1; -
 DR EMBL, U82664; AAB40164.1; -
 DR EMBL, AE005220; AAG54755.1; -
 DR EMBL, AP002551; BAB33882.1; -
 DR EMBL, S68715; AAC60469.2; -
 DR PIR, J00696; J00696.
 DR PIR, S12301; S12301.
 DR PIR, S12301; S12301.
 DR Ecogene; EG10938; *secD*.
 DR InterPro; IPR003335; *SecD_SecF*.
 DR Pfam; PF02355; *SecD_SecF_1*.
 DR TIGRfams; TIGR00916; 2A0504s01.1.
 DR TIGRfams; TIGR01129; *secD*.1.
 KW Protein transposer; Translocation; Transmembrane; Inner membrane;
 KW Complete proteome.
 FT DOMAIN 1
 FT TRANSMEM 1 9 CYTOPLASMIC (PROBABLE).
 FT DOMAIN 2 29 PROBABLE.
 FT TRANSMEM 30 455 PERIPLASMIC (PROBABLE).
 FT TRANSMEM 456 472 PROBABLE.
 FT DOMAIN 473 476 CYTOPLASMIC (PROBABLE).
 FT TRANSMEM 477 497 PROBABLE.
 FT DOMAIN 498 501 PERIPLASMIC (PROBABLE).
 FT TRANSMEM 502 518 CYTOPLASMIC (PROBABLE).
 FT DOMAIN 519 563 CYTOPLASMIC (PROBABLE).
 FT TRANSMEM 564 580 PROBABLE.
 FT TRANSMEM 581 585 PERIPLASMIC (PROBABLE).
 FT TRANSMEM 586 605 PROBABLE.
 FT TRANSMEM 606 615 CYTOPLASMIC (PROBABLE).
 FT DOMAIN 615 78 F -> S (IN REF. 1).
 FT CONFLICT 78 155 R -> A (IN REF. 1).
 FT CONFLICT 155 155
 SQ SEQUENCE 615 AA; 66632 MW; 1099EB6A9CC98EBE CRC64;

Alignment Scores:		
Pred. No.:	2,76e-21	Length: 615
Score:	359.50	Matches: 80
Percent Similarity:	62.39%	Conservative: 61
Best Local Similarity:	35.40%	Mismatches: 82
Query Match:	10.85%	Indels: 3
DB:	1	Gaps: 2
US-09-868-987-1 (1-1864) x SECD_ECOLI (1-615)		
QY	44 AATGACAAATATTTCTGCAAAACCGTGAGTGCGTATGATGATGACGTTATATG	1030
DB	375 SerGlyIysIysaSaPaIaAsn-----GlyArgAlaIaIeValIySgIngluIuVal	3922
QY	104 GTCAGACAGCCCTATTTTAAACGTCCTCCATTTGAAAATATCATGCCAGGTCTCAGGAAATTT	1633
DB	393 IleAenIleAlaAenIleAenIleGlnSerArgLeuGlyAAsnSerPheArgIleThrGlyIleAAsn	4121
QY	164 ACCCACCGCTGAAGTAGAGAAACTGCGCTCAGATTTAAATATCTGGAGGATGCTTTTGGT	2233
DB	413 AsnProAenGluAlaArgGlnLeuSerLeuLeuAArgAlaGlyAlaIeAPro	4322
QY	224 CCCGAGGTTCTCAGTGAAGAAGACCATCTCTTCGATCTTGGGAAAAAATCATGACACAA	2883
DB	433 IlegInIleValIgluIuAArgThrIleGlyProThrIleuGlyMetGlnAenIleGluGln	4532
QY	284 GGCATTATCTCAGCATGCGTGTGGCTTGGCAATGCTTATGTGTTGATGAGCGTATATAT	3433
DB	453 GlyLeuGluAlaIcySLeuAlaGlyLeuIeValSerIleuPheMetIleIlePheTyr	4727
QY	344 AGATTGAGAGCGGTCATCGCTTGGGAGCTGTCTCTTGATGATCTTTTGCTTATCTGGGCA	4033
DB	473 LysIysPheGlyLeuIleAAlaThrSerAlaLeuIleAlaAsnLeuIleIleValIglY	4923
QY	404 GCTCATCAGTATTG---GATGGCCCACTCACTGTCAGGACTCGCTGGGATTTGCTT	4603
DB	433 IleMetSerLeuLeuProGlyAlaIThrLeuSerMetProGlyIleAlaGlyIleValLeu	5122
QY	461 GCTATGGGATGCGCCGATAGATGACAAATGTTCTTGATTCGAAAGATCCAGAGGAATTT	5203
DB	513 ThrIleuAlaIaIaIaValaAspAlaAsnValIleuIleAenGluArgIleYsGluGluLeu	5323
QY	521 TTATGTCCTCAAGCTCTTAAAAAATCTGTAGAAAAAGATATACCAAGCTTTGGAGCC	5803
DB	533 SerAenGlyArgThrValGlnGlnAlaIleAspGluGlyTyrArgGlyAlaPheSerSer	5523
QY	581 ATTTTGAATTCGAACCTTGACTACAGTATGTGGCCGCACAGCTCTCTTCTCCATGATACA	6403
DB	553 IlePheAspAlaAsnIleThrIleuIleuIleYsValIleIleuIleuYrAlaValaGlyThr	5723
QY	641 GGGCCTATTAAGGTTTGGCTTTCGACATTGATTTTAGAAATTTCTCTCAATGTTTACG	7003
DB	573 GlyAlaIleIySgIyPheAlaIleIleThrThrGlyIleGlyAlaAlaThrSerMetPheThr	5923
QY	701 GCTCTTTTCAGACTAA 718	
DB	593 AlaIleValGlyThrArg 598	
RESULT 14		
SECD_SALCH	STANDARD;	PRT; 615 AA.
ID_SECD_SALCH	Q92FF8;	
AC	30-MAY-2000 (Rel. 39, Created)	
DT	30-MAY-2000 (Rel. 39, Last sequence update)	
DT	30-MAY-2000 (Rel. 39, Last annotation update)	
DE	Protein-export membrane protein secd.	
GN	SECD.	
OS	Salmonella cholerae-suis (Salmonella enterica).	
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;	
OC	Bacteria; Salmonella.	
OX	NCBI_TaxID=591;	
RN	(1)	
RP	SEQUENCE FROM N.A.	

```

533 SerAsnGlyArgThrValGlnGlnAlaIleAsnGluGlyTyrAlaGlyAlaPheSer 552
581 ATTTTGTATTAACTTGACTCAGATATTGGCTCAGCACATCTTTTCTTCCTAGATACA 640
553 IlePheAspAlaAsnIleThrThrLeuIleValIleIleLeuTyrAlaValGlyThr 572
641 GGGCCATTAAAGGGTTTCCTTGACATGATTATTAGGAATTTCTCTCAATGTTTACG 700
573 GlyAlaIleGlyGlyPheAlaIleThrThrGlyIleGlyValAlaThrSerMetPheThr 592
701 GCTCTTTTCATGACTAAA 718
593 AlaIleIleGlyThrArg 598

RESULT 15
SECF_HELPJ
ID SECF_HELPJ STANDARD; PRT; 323 AA.
AC Q9ZJ65;
DC 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DD 16-OCT-2001 (Rel. 40, Last annotation update)
DE Protein-export membrane protein secF.
GN SECF OR JHP1450
OS Helicobacter pylori J99 (Campylobacter pylori J99).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
NCBI TaxID=85963;
[1]_TaxID=85963;
SEQUENCE FROM N.A.
MEDLINE=99120557; PubMed=9923682;
Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,
Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,
Tummino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,
Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,
Trust T.J.;
"Genomic sequence comparison of two unrelated isolates of the human
gastric pathogen Helicobacter pylori.";
Nature 397:176-180(1999).
-!- FUNCTION: INVOLVED IN PROTEIN EXPORT (BY SIMILARITY).
-!- SUBUNIT: PART OF THE PROKARYOTIC PROTEIN TRANSLOCATION APPARATUS
WHICH COMPRISE SECA, SECB, SECD, SECE, SECF, SECG AND SECY
(BY SIMILARITY).
-!- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).
-!- SIMILARITY: BELONGS TO THE SECD/SECF FAMILY. SECF FAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation on
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; AE001567; AACD07023.1; -.
InterPro; IPR003335; SecD SecF.
Pfam; PF02355; SecD SecF; 1.
TIGRFAMs; TIGR009167; 2A0604801; 1.
TIGRFAMs; TIGR00966; 3A0501807; 1.
Protein transport; Translocation; Transmembrane; Membrane;
Complete proteome.
KW TRANSMEM 19 39 POTENTIAL.
FT TRANSMEM 138 158 POTENTIAL.
FT TRANSMEM 162 182 POTENTIAL.
FT TRANSMEM 244 264 POTENTIAL.
FT TRANSMEM 269 289 POTENTIAL.
SEQUENCE 323 AA; 36142 MW; 15363AA8FF0E9215 CRC64;

Alignment Scores:
Pred. No.: 1,24e-20 Length: 323
Score: 351.00 Matches: 91
Percent Similarity: 50.81% Conservative: 66
Best Local Similarity: 29.45% Mismatches: 104

```

```

Query Match:      10.60%      Index:      48
DB:               1          Gaps:         7
US-09-866-987-1 (1-1864) x SECF_HELPJ (1-323)

```

[illegible]

```

DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Protein-export membrane protein SecF.
GN SEC_F OR HP1549.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_Textid=210;
XN [1]
XP SEQUENCE FROM N.A.
RC STRAIN=26695 / ATCC 700392;
RX MEDLINE=97394467; PubMed=9252185;
RA Tomb J.-F., White O., Kellavage A.R., Clayton R.A., Sutton G.G.,
RA Fleischmann R.D., Ketchum K.A., Klein H.-P., Gill S., Dougherty B.A.,
RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
RA Loftus B., Richardson D., Dodson R., Khalak H.G., Glocke A.,
RA McMeney K., FitzGerald L.M., Lee N., Adams M.D., Hickey E.K.,
RA Berg D.E., Gocayne J.D., Uiterback T.R., Peterson J.D., Kelley J.M.,
RA Cotton M.D., Meldrum J.M., Fujii C., Bowman C., Matthey L., Wallin E.,
RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
RA Venter J.C.;
RT "The complete genome sequence of the gastric pathogen Helicobacter
RT pylori."
RL Nature 388:539-547(1997).
CC -1- FUNCTION: INVOLVED IN PROTEIN EXPORT (BY SIMILARITY).
CC -1- SUBUNIT: PART OF THE PROCAROTIC PROTEIN TRANSLOCATION APPARATUS
CC WHICH COMPRISE SEC_A, SEC_B, SEC_C, SEC_D, SEC_E, SEC_F, SEC_G AND SEC_H
CC (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).
CC -1- SIMILARITY: BELONGS TO THE SEC/SEC_F FAMILY. SEC_F FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collabor
CC between the Swiss Institute of Bioinformatics and the EMBL outstat
CC the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no
CC modified and this statement is not removed. Usage by and for commec
CC entities requires a license agreement (See http://www.isb-sib.ch/anno
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE000652; AAD08587.1; -.
CC DR TIGR; HP1549; -.
CC DR InterPro; IPR003335; SecF SecF.
CC DR Pfam; PF02355; SecF SecF.1.
CC DR TIGRFAMs; TIGR009167.2A0604801; 1.
CC DR TIGRFAMs; TIGR00966; 3a501807; 1.
CC KW Protein transport; Translocation; Transmembrane; Membrane;
CC Complete proteome.
CC FT TRANSMEM 19 39 POTENTIAL.
CC FT TRANSMEM 138 158 POTENTIAL.
CC FT TRANSMEM 162 182 POTENTIAL.
CC FT TRANSMEM 244 264 POTENTIAL.
CC FT TRANSMEM 269 289 POTENTIAL.
CC SQ SEQUENCE 323 AA; 36276 MW; EAD9190473F21F59 CRC64;

```

QY 950 ATCAGCGATGTTGCTCAATCGTGGC-----AAAGTTGTGCATAAA 991
 Db 49 -----SerLeuValGlnValArgThrGlnAenAlaProIleLysGluValArgAsp 66
 QY 992 CTACAGGAAGCTGGTCTTCTCTAGAGACTTCGATTCACAACTTTGGATCTTCAGAA 1051
 Db 67 LeuPheGluLysGluAlaArgPheLysGlyValGlnValSerGluPheGlySerLysGlu 86
 QY 1052 AAGATCAAAATCTATTATTAGTATAAAGCTTTAAGCTTACTACTACGACGACGCTCT 1111
 Db 87 GluIleLeuLysPheProPheValGluThrAlaGluAenGluAenGluAenAla--- 105
 QY 1112 CTCCTAAATTAACGATCATGAGCTGGGCTTATTGTGGATTGTGTGAGAAACAGGCT 1171
 Db 106 -----IleValAlaAen----- 109
 QY 1172 AGATTCTCTACGGAACCTCTAAACGAAACGCAAAATTTTGGTCAAGGTAAAGCAGCAA 1231
 Db 110 ---IleLeuLysProSerGlyAspPheGluLysPheAspThrValGlyProArg 128
 QY 1232 CTATCGAAGAAATGCGTTATCAGCGACCATCGGCTTTTAGGAGCTTTTGGCAATCATC 1291
 Db 129 ValGlySerGluLysGluLysGlyLeuSerLeuLysLeuAlaLeuAlaIle 148
 QY 1292 TTGCTCTATGAGTTGCGCTTTGAATGGCAATATGCTTTTCAGTGGCGTATGCGCTTTA 1351
 Db 149 MetValTyValSerPheArgTyGluTrpArgPheAlaLeuAlaSerValIleAlaLeu 168
 QY 1352 ATTCATGAC-----CTTTGCGTACCTGTCAGCTCTGTTTATAGCACATTTCTTTTG 1405
 Db 169 ValHisAspValIleLeuValAlaSerSerValIleValPhe----- 182
 QY 1406 AAGAAATTCAAATAGATTTCAGGCCATTCGTTTAAATGACTGTATTGGGTATTTC 1465
 Db 183 ---LysIleAspMetAenLeuGluValIleAlaLeuLeuThrLeuLysGlyTySer 201
 QY 1466 TTAACAATATCTGATCATTTTGTATGCTGTTTGTGAAGATCGCAAGCAACTGTTT 1525
 Db 202 IleAenAspThrIleIlePheAspArgIleArgGluGluMetLeuSerGlnLysThr 221
 QY 1526 ACCCTATGATGTTTATAGTTAATGATCCCTTCAAAAGAGCTTCAGCGCAGCGTAATG 1585
 Db 222 LysAenAlaThrGlnAlaIleAspGluAlaIleSerSerThrLeuThrArgThrLeuLeu 241
 QY 1586 ACAACAGCTACAACCTATCATGTTTGTATGCTGTTTGTATAGCGGCTCCTCTGTC 1645
 Db 242 ThrSerLeuThrValPhePheValValLeuLeuLeuCysValPheGlySerLysIleIle 261
 QY 1646 TTTAATTTGCTATTATATGACCATAGGATTTCTTCTAGGAATTTATGCTCTCTTAT 1705
 Db 262 IleGlyPheSerLeuProMetLeuIleGlyThrIleValGlyThrTySerSerIlePhe 281
 QY 1706 ATTCCACCACCTCTGTTGTTGTTATG 1732
 Db 282 IleAlaProLysValAlaLeuLeuLeu 290

RESULT 17
 SECFC_RICPR
 ID SECFC_RICPR STANDARD; PRT: 311 AA.
 AC Q2E34;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Protein-export membrane protein secf.
 GN SECFC OR RP14.
 OS Rickettsia prowazekii.
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
 OC Rickettsiaceae; Rickettsia;
 OX NCBI_TaxID=782;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Madrid E;
 RX MEDLINE=99039499; PubMed=9823893;

RA Andersson S.G.E., Zomorodipour A., Andersson J.O.,
 RA Scheritz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,
 RA Eriksson A.-S., Winkler H.H., Kurland C.G.;
 RT "The genome sequence of Rickettsia prowazekii and the origin of
 RT mitochondria".
 RL Nature 396:133-140(1998).
 CC -1- FUNCTION: INVOLVED IN PROTEIN EXPORT.
 CC -1- SUBUNIT: PART OF THE PROKARYOTIC PROTEIN TRANSLLOCATION APPARATUS
 CC WHICH COMPRISE SECA, SECB, SECD, SECE, SECF, SECG AND SECH
 CC (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).
 CC -1- SIMILARITY: BELONGS TO THE SECY/SECF FAMILY. SECY FAMILY.
 CC
 CC THIS SWISS-PROT entry is copyrighted. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: AJ235270; CAA14583.1; -
 DR InterPro: IPR003335; SecD_SecF.
 DR Pfam: PF02355; SecD_SecF.1.
 DR TIGRFAMs: TIGR00916; 2A0604801; 1.
 DR TIGRFAMs: TIGR00966; 3a0501807; 1.
 KW Protein transport; Translocation; Transmembrane; Membrane;
 KW Complete proteome.
 FT TRANSMEM 36 56 POTENTIAL.
 FT TRANSMEM 140 160 POTENTIAL.
 FT TRANSMEM 164 184 POTENTIAL.
 FT TRANSMEM 194 214 POTENTIAL.
 FT TRANSMEM 246 266 POTENTIAL.
 FT TRANSMEM 272 292 POTENTIAL.
 SQ SEQUENCE 311 AA; 35175 MW; 31863A150F32E630 CRC64;

Alignment Scores:
 Pred. No.: 6,58e-20 Length: 311
 Score: 342.00 Matches: 94
 Percent Similarity: 49.85% Conservative: 72
 Best Local Similarity: 28.23% Mismatches: 113
 Query Match: 10.33% Indels: 54
 DB: 1 Gaps: 13

US-09-868-987-1 (1-1864) x SECFC_RICPR (1-311)

QY 788 ATAAAGCATGATCTCTGAGAGGATGCAAAAACCTTTGGCTGTTCTGGAAGTGTCTT 847
 Db 13 IleAspPheAspPheMetAenPheLysLysValSerTySerPheSer-----Ile 29
 QY 848 CTTTGTAGTTGCGTTGCTCGGTTTGGAGCTGGAATTCGGT-----TTG 895
 Db 30 IleLeuSerLeuLeuSerLeu-----IleTrpIleSerIleTyLysPheAsnPhe 46
 QY 896 GGAATGGAATTTAAAGAGGATGATGCTTTACCTTTAAT----- 934
 Db 47 GlyIleAspPheValGlyIleValIleGluValArgLeuAspGlnAlaProAspLeu 66
 QY 935 CCAAAAGAGCATGCGATCAGCATGTTGCTCAAAATGCGTGGCAAAAGTTGTGCATAACTA 994
 Db 67 ProllysMetArgAlaValLeuSerAlaLeuGluIle---GlyGluValVal-----Leu 83
 QY 995 CAGGAAGCTGCTTTCTTCTAGAGACTTCGGTATTCAACATTTGGATCTTCAGAAAG 1054
 Db 84 GluAenPheGly---SerGluArgAspLeuSerIleArg---PheGlySerSerSerGlu 101
 QY 1055 ATCAAAATCTATTATTAGTATAAAGCTTTAAGCTTATAGCTATAGCAGATCAGGCTCTCTC 1114
 Db 102 -----GluAenLeuMetLysAenIleAspIleIleLysThrSerLeu 115
 QY 1115 CTAAATTAACGATCATGAGCTGGCGTTATTGTGGGATTGTTGTAGAAACAGCGCTAGA 1174
 Db 116 -----ArgAsnAen 118

QY 1175 TTCTCTACGGAACCTTAAGCAAAATTTGTCAAGGTACGCAAACTA 1234
 DB 119 PheProlyraenPheGlyTyrArgLysValasphe-----ValGlyProGlnVal 135
 QY 1235 TCGAAGAAATGCGTATCATCGGCGACATCGGCTTTTGAAGCTTTGGCATCATCTTG 1294
 DB 136 GlyArgGlnLeuIleGlnAlaGlyAlaMetLeuMetLeuPheSerPheLeuAlaIleMet 155
 QY 1295 CTCTATGTGAGTTGGCTTGGTGAATGCAATATGCTTTCAGTGGCGCTTAAAT 1354
 DB 156 ValTyrIleGlyValArgPheGlnTyrPheGlyIleLeuIleAlaLeuVal 175
 QY 1355 CATGACCTTTGGCTACCTGCGACCTCTGTTTATGACATTTCTTTGAGAAAT 1414
 DB 176 HisaspVal-----IleLeuAlaLeuGlyPheMetSerMetThrLysLeu 190
 QY 1415 CAATATGATTTGCAAGCATTTGCTTAAATGACTGATTTGGGATTTAAACAT 1474
 DB 191 AspheAsnLeuSerThrIleAlaAlaValLeuThrIleIleGlyTyrSerValAsnAsp 210
 QY 1475 ACTTGATCATTTTGTATGCTGATTCGTAAGATCGCAAGCAACCTTTAACCCATG 1534
 DB 211 SerValValIleTyrAspArgIleArgGlnAsnLeuArgLysTyrHisLysAsnIle 230
 QY 1535 CATGTTTATGTAATGATGCCCTTCAAAAGACCTTCAGCCGACGCTAATGACACACT 1594
 DB 231 ThrGlnIleIleAsnLeuSerIleAsnGlnIleuThrLeuSerArgThrIleLeuThrValIle 250
 QY 1595 ACAACTATCATGTTTGTATGCTTGTATGCTTGTATGCTGCTGCTTAAATTT 1654
 DB 251 ThrThrLeuLeuAlaAsnLeuAlaLeuLeuLeuPheGlyGlyAlaIleHisSerPhe 270
 QY 1655 GCATTTATATGACATAGAGATGATCTTCTAGCACTTATGCTCTTATATATGACCA 1714
 DB 271 SerValLeuValAlaPhePheGlyIleIleAlaGlyTyrSerSerIlePheIleSerAla 290
 QY 1715 CCTCGTTG---TTGTTATGCTCGCTTAAAGAAATCGC 1750
 DB 291 ProIleuThrMetPheAlaAsnArgLysPheAsnLys 303
 RESULT 18
 SECF_HAEIN STANDARD; PRT; 325 AA.
 AC P44590;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Protein-export membrane protein secf.
 GN SECF OR HI0239.
 OS Haemophilus influenzae.
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
 OC Haemophilus.
 CX NCBI_TaxID=727;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Rd / KW20 / ATCC 51907;
 RX MEDLINE=95350630; PubMed=7542800;
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
 RA Ketchum K.A., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
 RA McEweney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
 RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
 RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
 RA Uetrich T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
 RA Fine L.D., Fritchman J.L., Fuhmann J.L., Geoghegan N.S.M.,
 RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
 RA Venter J.C.;
 RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
 Rd.";
 RL Science 269:496-512(1995).
 CC -1- FUNCTION: INVOLVED IN PROTEIN EXPORT (BY SIMILARITY).
 CC -1- SUBUNIT: PART OF THE PROKARYOTIC PROTEIN TRANSLOCATION APPARATUS
 CC WHICH COMPRISE SECA, SECB, SECD, SECE, SECF, SECG AND SECH

CC (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
 CC (By similarity).
 CC -1- SIMILARITY: BELONGS TO THE SECD/SECF FAMILY. SECF FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U32710; AAC21907.1; -.
 DR TIGR: HI0239; -.
 DR InterPro: IPR003335; SecD_SecF.
 DR Pfam: PF02355; SecD_SecF.1.
 DR TIGRFAMs: TIGR00916; 2A0604s01; 1.
 DR TIGRFAMs: TIGR00966; 3a0501s07; 1.
 KW Protein transport; Translocation; Transmembrane; Inner membrane;
 KW Complete proteome.
 FT TRANSMEM 36 56 POTENTIAL.
 FT TRANSMEM 148 168 POTENTIAL.
 FT TRANSMEM 187 207 POTENTIAL.
 FT TRANSMEM 254 274 POTENTIAL.
 FT TRANSMEM 281 301 POTENTIAL.
 SO SEQUENCE 325 AA; 36051 MW; BRC59239EDICF1F2 CRC64;
 Alignment Scores:
 Pred. No.: 6,62e-20 Length: 325
 Score: 342.00 Matches: 98
 Percent Similarity: 51.90% Conservative: 80
 Best Local Similarity: 28.57% Mismatches: 111
 Query Match: 10.33% Indels: 54
 DB: 1 Gaps: 15
 US-09-868-987-1 (1-1864) x SECF_HAEIN (1-325)
 QY 710 ATGACTAATTTTCTTCATGCTGTGATGATTAAGCCCACTACAGTTCATATG 769
 DB 1 MetMetLysLeuPhe-----ThrLysAspLysAspGlyHisPhe 13
 QY 770 ATGATATAGTTCTGCGGATTAAG-----CATGATTTCTTGAAGATGC 814
 DB 14 IleArgIleLeuAsnGlyIleLysLeuProPheProLeuThrGlnPheMetLysValArg 33
 QY 815 AAAAACTTTGGCGCTGTTTGTGAAGTGTTCCTTTTACGT-----TGGCTTGT 865
 DB 34 LysLeuGlyTyrIleLeuSerAlaLeuMetValIleSerLeuPhePheIleThr 53
 QY 866 CTGCGTTTGGAGCTTGAATTCGTTTGGGAATGATTTTAAAGAGGATACCTT 925
 DB 54 LysGlyPheAsn---TTP-----GlyLeuAspPheThrGlyValAlaPhe 68
 QY 926 ACCTTATATCAAAAGATGATGATGATGATGATGATGATGATGATGATGATG 985
 DB 69 AsphThrHisPheSerGlnSer-----AlaAsnLeuGlnIleArgSer----- 83
 QY 986 CATTAATACAGAGAGAGCTGCTTCTTCTTGAAGACTTCGATTAACAATTTGATCT 1045
 DB 84 ---LysLeuHisGlnAlaGlnGlyIleGlnSerPro-----IleValGlnThrThrGlySer 100
 QY 1046 TCAGAAAAGATCAAAATCTATTTTATGATTAAGCTTTAAGCTTATACAGATACGA 1105
 DB 101 ValGlnAspValMetIle-----ArgLeuPro 109
 QY 1106 GCTCTCTCTAAATTAATGATCATGATGCTGGGATATGTTGATGATGATGATG 1165
 DB 110 AlaSerAsnAsnAspSerThrIle-----GlyGlnHisValLysSerMetLeuGlnAsn 127
 QY 1166 AGGCTAGATTTCTCTACGGAATCTTAACGAAGCAAAATTTTGGTCAAAAGTAA 1225
 DB 128 ValAspLysAspIle-----GlnIleArgSerIleGlnPhe-----ValGly 141


```

Db 538 GlyhehaIatPseSerLeuSerValGlyIleValAlaSerLeuPheSerSerLeuIlePhe 557
Qy 713 ACTAAATTTTC-----TTCATGCTGTGATGAATTAAGACCAACATACAGTTGCAT 766
Db 558 SerArgPheIleLeuGluPheIleIleSerValArgLysSerLysPheIleSerIleSer 577
Qy 767 ATGATGATTAAGTTC 781
Db 578 TtpSerSerLysTyr 582

RESULT 20
SECF_AQUAE STANDARD: PRT: 288 AA.
AC 067536;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Protein-export membrane protein secf.
GN SECF OR AQ 1602.
OS Aquifex aeolicus.
OC Bacteria; Aquificae; Aquificae (class); Aquificales; Aquificaceae;
OC Aquifex.
OX NCBI_TaxID=63363;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VF5;
RX MEDLINE=98196666; PubMed=9537320;
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA Graham D.E., Overbeek R., Sneed M.A., Keller M., Aubay M., Huber R.,
RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT "The complete genome of the hyperthermophilic bacterium Aquifex
RT aeolicus."
RL Nature 352:353-358(1998).
CC -1- FUNCTION: INVOLVED IN PROTEIN EXPORT (BY SIMILARITY).
CC -1- SUBUNIT: PART OF THE PROKARYOTIC PROTEIN TRANSLLOCATION APPARATUS
CC WHICH COMPRISE SECA, SECB, SECD, SECE, SECF, SECG AND SECY
CC (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (by similarity).
CC -1- SIMILARITY: BELONGS TO THE SECD/SECF FAMILY. SECF FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: AE000747; AAC07496.1; -
DR InterPro: IPR003335; SecD_SecF.
DR Pfam: PF02355; SecD_SecF.1.
DR TIGRFAMs: TIGR00916; 2A0604s01; 1.
DR TIGRFAMs: TIGR00966; 3A0501s07; 1.
KM Protein transport; Translocation; Transmembrane; Complete proteome.
FT TRANSMEM 13 33 POTENTIAL.
FT TRANSMEM 123 143 POTENTIAL.
FT TRANSMEM 145 165 POTENTIAL.
FT TRANSMEM 175 195 POTENTIAL.
FT TRANSMEM 240 260 POTENTIAL.
SQ SEQUENCE 288 AA; 31632 MW; E526B8D6BBD99770 CRC64;

Alignment Scores:
Pred. No.: 1.42e-17 Length: 288
Score: 313.00 Matches: 93
Percent Similarity: 51.08% Conservative: 72
Best Local Similarity: 28.79% Mismatches: 110
Query Match: 9.45% Indels: 48
DB: 1 Gaps: 12

US-09-868-987-1 (1-1864) x SECF_AQUAE (1-288)
Qy 800 TTCTTGAAGATGCAAAACTTTGGGCTTTTCTGGAAGTGTTCCTTTAGGTTGC 859
|||||
::: |||||
::: |||||

```

```

Db 6 PheLeuArgLeuArgLysAlaIleTyrGlyValSerAlaLeuLeuIleSerLeu 25
Qy 860 GTTGCTCTCGGGTTTGGAGCCTCGTAATCCGTTTGGGAATGATTTTAAAGAGG--- 916
Db 26 LeuSerLeuLeuPheArgGlyLeuAsn-----LeuGlyLeuAspPheThrGlyGlyThr 43
Qy 917 ---TATGCTTTACCTTTATTCAGAAAGAGCATGCGATCGCATGTGCTCAATCGCT 973
Db 44 LeuTyrGluValLysPheGlyLysSer-----ValAspIle----- 55
Qy 974 GCGAAGTGTGCACTAACTACAGAGAGCTGTCTTCTCTAGACCTTCGATTAAC 1033
Db 56 GlyLysLeuArgLysThrLysSerIleGlyIle-----LysGlyPheLeuIleGln 73
Qy 1034 ---ACATTGCAATCTTCAGAAAGATCAAAATCTATTATTAAGTAAGCTTAAGCTAT 1090
Db 74 GluThrLysGluGlyThrPheValIleLysVal-----LysThrGlyGluPro 89
Qy 1091 ACTAAGCAGATAGAGGCTCTCTCTAAATTAACGATCATGAGCTGCGCTTATGTGGG 1150
Db 90 ValGluLysLeuGluLysPheValLysLys-----Val----- 99
Qy 1151 ATTGTTGTCAAGAAACAGGCTTAGATTCTCTACAGAAACTCTAAGCAAGCAAAATTT 1210
Db 100 -----PheGlyLysTyrGluLeuIleArgLys--- 108
Qy 1211 TGGTCAAGATTAAGCAAGCAACTATGCAAGAAATGCGTTATGAGGACCATCGGCTT 1270
Db 109 ---GluThrLysSerGlyValLysSerGluGluLeuGlnLysAlaValPheAlaIle 127
Qy 1271 TTGAGACCTTGGCAATCACTTGTCTATGATGAGTGTGCTTGAAGTGAATGCT 1330
Db 128 LeuThrAlaLeuGlyLysIleLeuLeuTyrLysGlyValArgPheGlnProValIleGly 147
Qy 1331 TTGAGTCCGTATGCGCTTAACTATGACCTTTGGCTACCTGTGACGCTTGTATTA 1390
Db 148 PheGlyValAlaLeuAlaLeuAlaHisAspValIleThrVal-----Leu 152
Qy 1391 GCACATTTCTTTTGAAGAAATCAATATGATTTGCAAGCCATTGGTCTTAAATGACT 1450
Db 163 GlyAlaTyrSerIleThrGlnArgGluValAsnLeuGluValAlaSerAlaIleLeuVal 182
Qy 1451 GTATGGGATTCATTAACATCACTTGAATGATTTGATTTGATTTGATTTGATTTGATTT 1507
Db 183 ValAlaGlyLysSerValAlaAspThrValIlePheAspArgLysGluAsnLeu 202
Qy 1508 CGCCAAAGCAAGCTGTTTACCTTATGATGATTTGATTTGATTTGATTTGATTTGATTT 1567
Db 203 ArgLysLysLysGlyPheThr-----LeuGluGluIleMetAspLeuSerIleAsnGlnThr 221
Qy 1568 TTACGCGCAGCGTATGACACAGCTACATCTTATGATTTGATTTGATTTGATTTGATTT 1627
Db 222 LeuSerArgThrIleMetThrSerLeuThrThrLeuValThrAlaLeuThrLeuPheIle 241
Qy 1628 ATGAGGGGCGCTCTGCTTATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTT 1687
Db 242 LeuGlyGlyLysThrAlaLeuSerAsnIleMetPheAlaPheValValGlyValGly 261
Qy 1688 ACTTATGCTCTCTTATATGACACACCTCTGTTGTTTATGCTGCTGCTGCTGCTGCTGCT 1747
Db 262 ThrTyrSerSerValPheValAlaSerAlaPheValLeuAspMetGlnLysLeuPheLys 281
Qy 1748 CGCTCAAA 1756
Db 282 ArgGlyGlu 284

RESULT 21
SECD_MYCLE STANDARD: PRT: 571 AA.
AC P38387; O69485; Q9CCT3;
DT 01-OCT-1994 (Rel. 30, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)

```

DE Protein-export membrane protein secD.
GN SEC D OR ML0487 OR MLCB1259.05 OR B1177_C1_164.
OS Mycobacterium leprae.
OC Bacteria; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TN;
RX MEDLINE=21128732; PubMed=11234002;
RA Cole S.T., Eiglmier K., Parkhill J., James K.D., Thomson N.R.,
RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R.M., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA Barrall B.G.;
RT "Massive gene decay in the leprosy bacillus."
RL Nature 409:1007-1011(2001).
RN [2]
RP SEQUENCE OF 88-571 FROM N.A.
RA Smith D.R., Robison K.;
RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: INVOLVED IN PROTEIN EXPORT (BY SIMILARITY).
CC -!- SUBUNIT: PART OF THE PROKARYOTIC PROTEIN TRANSLOCATION APPARATUS
CC WHICH COMPRISE SECA, SECB, SECD, SECE, SECF, SECG AND SECH
CC (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).
CC -!- SIMILARITY: BELONGS TO THE SECD/SECF FAMILY. SECD FAMILY.
CC -!- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN FROM POSITION 567
CC ONWARD DUE TO A FRAMESHIFT.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AL023591; CAAL0980.1; -
CC EMBL; AL583918; CAC29995.1; ALT INIT.
CC EMBL; U00011; AAA17086.1; ALT_FRAME.
CC Leproma; ML0487; -
CC InterPro; IPR003335; SecD SecF.
CC TIGRFAMs; TIGR00916; 2A0604501; 1.
CC TIGRFAMs; TIGR01129; secD; 1.
CC Protein transport; Translocation; Transmembrane; Membrane;
CC Complete proteome.
CC TRANSMEM 13 33 POTENTIAL.
CC TRANSMEM 383 403 POTENTIAL.
CC TRANSMEM 408 428 POTENTIAL.
CC TRANSMEM 439 459 POTENTIAL.
CC TRANSMEM 487 507 POTENTIAL.
CC TRANSMEM 512 532 POTENTIAL.
CC SEQUENCE 571 AA; 60718 MW; 40852F70B049B496 CRC64;

Alignment Scores:
Pred. No.: 8.16e-17 Length: 571
Score: 304.00 Matches: 77
Percent Similarity: 51.52% Conservative: 59
Best Local Similarity: 29.17% Mismatches: 102
Query Match: 9.18% Indels: 26
DB: 1 Gaps: 5

US-09-868-987-1 (1-1864) x SECD_MYLE (1-571)

OY 5 ACTTCGGCATATGTTCAGAGGGGATCAGCGCACTGCTAATGGACATAT----- 55
Db 275 ThrSerGlyMetAsnGlnArgGlyValGlyTyrValValAspLeuGlnPheLysSerAla 294

OY 56 -----TCTGCAACCGTGGATGGCGTATGGCTGTAGTG 88
Db 295 AlaAlaAspValTrpAlaAspPheThrAlaAlaHisIleGlyThrGlnThrAlaPheThr 314
OY 89 ATTGACGGTTATATGTCAGCAGCCCTATTAAAC---GTCCCATTTGAAAAAATCATGCC 145
Db 315 LeuAspSerGluValValSerValProValIleAsnGluAlaIleLeuGlyGlyArgThr 334
OY 146 AGTGTCTCAGGG-----AAATTTACCACCGTGAAGTGAGCAAACTCGCCTCAGAT 196
Db 335 GlnIleSerGlyGlyAspProPheThrAlaAlaThrAlaArgGlnLeuAlaAsnVal 354
OY 197 TTAATACTCGAGGATGCTCTTTGTTCCGAGGTTCTCAGTGAAGAGACATCTCTTCT 256
Db 355 LeuIlyTyrGlySerLeuProLeuSerPheGluProSerGluAlaGlnThrValSerAla 374
OY 257 GATCTCGGAAAAACAATGTACCAAGGATATTCTCAGCATGCTGTGGCTTCGCAATG 316
Db 375 ThrLeuGlyLeuThrSerLeuArgAlaGlyLeuIleAlaGlyAlaIleGlyLeuSerLeu 394
OY 317 CTATATGTTTGTATGAGCGTATATTATAGATTTTGGAGCGCTCATCGCTTCGGGAGCTGTT 376
Db 395 ValLeuLeuTyrSerLeuLeuTyrTyrArgValLeuGlyLeuLeuThrAlaPheSerLeu 414
OY 377 CTCTCGAATCTTTGCTTATCTGGGCGAGCTCTA-----CAGTATTTGGATGCG 424
Db 415 PheCysSerGlyThrIleIlePheAlaIleLeuValLeuLeuGlyArgTyrIleAsnTyr 434
OY 425 CCACCTACCTGTGTCAGACTCGCTGGGATGTTCTTCTGCTATGGGATGGCGGTAGATGCA 484
Db 435 ThrLeuAspLeuAlaGlyIleAlaGlyLeuIleGlyIleGlyThrThrAlaAspSer 454
OY 485 AATGTTCTTGTTATCGAAAGAAATCCGAGAGAAATTTTATTTGTTCTCAAGTCTTAAAAA 544
Db 455 PheValValPhePheGluArgIleIleAspGluIleArgGluGlyArgSerPheArgSer 474
OY 545 TCTGTAGAAAAGATATACCAAGGCTTTTGGAGCCATTTTGTGATTTCTAACTGACTACA 604
Db 475 AlaValProArgGlyTyrValArgAlaArgIleValSerGlyAsnAlaValThr 494
OY 605 GTATTGCGCTCAGCACTTCTTTCTCTAGATACAGGGCTATTAAAGGTTTTCITTG 664
Db 495 PheLeuAlaAlaValLeuHisPheLeuAlaIleGlyGlnValIleGlyPheAlaPhe 514
OY 665 ACATTGATTTTAGGAATTTCTCTCAATGTTTACGGCTCTTTTCATGCAATAATTTTC 724
Db 515 ThrLeu-----GlyLeuThrThrLeuAspLeuValValPhe 528
OY 725 TTCATGCTGTGG 736
Db 529 LeuValThrTrp 532

RESULT 22
SECD MYCTU STANDARD; PRT; 573 AA.
AC Q50634;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Protein-export membrane protein secD.
GN SEC D OR RV2587C OR MT2664 OR MTCY227.14.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae (class); Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37Rv;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Baeham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holtroyd S.,

RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
 RA Salston J.E., Taylor K., Whitehead S., Barrett B.G.;
 RT "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence";
 RL Nature 393:537-544(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CDC 1551 / Oshkosh;
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 RA Peterson J., Deboy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
 RA Kolonay J.F., Nelson W.C., Umayan L.A., Ermolaeva M.D., Salzberg S.L.,
 RA Delcher A., Utterback T., Weidman J., Kouri H., Gill J., Mikula A.,
 RA Bhalai W.;
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
 RT laboratory strains";
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: INVOLVED IN PROTEIN EXPORT.
 CC -1- SUBUNIT: PART OF THE PROKARYOTIC PROTEIN TRANSLLOCATION APPARATUS
 CC WHICH COMPRISE SECF, SECB, SECD, SECE, SECF, SECG AND SECY
 CC (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).
 CC -1- SIMILARITY: BELONGS TO THE SECD/SECF FAMILY. SECD FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements>
 CC or send an email to license@ebi.ac.uk).
 CC -----
 DR EMBL: Z77724, CAB01256.1;
 DR EMBL: AE007100, AAK46977.1; ALT_INIT.
 DR TIGR: MT2664;
 DR Tuberculist; Rv2587c;
 DR InterPro: IPR00335; SecD_SecF.
 DR Pfam: PF02355; SecD_SecF; 1.
 DR TIGRFAMs: TIGR00916; 2A064s01; 1.
 DR TIGRFAMs: TIGR01129; secD; 1.
 KM Protein transport; Translocation; Transmembrane; Membrane;
 KW Complete proteome.
 FT TRANSMEM 13
 FT TRANSMEM 385 405 POTENTIAL.
 FT TRANSMEM 410 430 POTENTIAL.
 FT TRANSMEM 441 461 POTENTIAL.
 FT TRANSMEM 489 509 POTENTIAL.
 FT TRANSMEM 514 534 POTENTIAL.
 SQ SEQUENCE 573 AA; 60:67 MW; 5A8F42EC0389CE2 CRC64;
 Alignment Scores:
 Pred. No.: 9.83e-17 Length: 573
 Score: 303.00 Matches: 78
 Percent Similarity: 51.13% Conservative: 58
 Best Local Similarity: 29.12% Mismatches: 100
 Query Match: 9.15% Indels: 30
 DB: 1 Gaps: 6
 US-09-868-987-1 (1-1864) x SECD_MYCTU (1-573)
 QY 5 ACTTCGCAATTTGTCAGAGGAGTCAGCGGCACTGCTAATGCAATAT----- 55
 DB 277 ThirsergiymetaspGlnargIlylelyrValValaspLeuGlnphelyselyPro 296
 QY 56 -----CTGCAACCGTGGATGGCGATGGCGATGGCTAGTG 88
 DB 297 AlaalaanileTtpalAspTyrThralaahsileglymrglnThralahethr 316
 QY 89 ATTGACGGTTATGTTAGAGAGCCCTATTAAAGTCCTCATGAA---AATCATGCC 145
 DB 317 LeuaspserGlnvalValserAlaahproGlnleGlnAlaalleProGlyGlyArgThr 336

QY 146 AGTGTCTCAGAG-----AAATTACCCAGCGTAAGTACGAACAACTCGCTCAGAT 196
 DB 337 GlnleaserIlylaspProPophethrAlaThralaargGlnleuAlaahVal 356
 QY 197 TTAATCTGAGAGATGTCCTTTTGGTCCGAGTTCTCAGTAAGAGATCTCTTC 256
 DB 357 LeuIytrIlyserleuProlseuserPheGlnProserGlnIlyGlnThrala 376
 QY 257 GATCTTGGGAAAAAACAATATACACAGGATATCTCAGCATGCTGGCTGGCATG 316
 DB 377 ThirleuIlyleuSerleuArgalaglymcelAlaIlyleuLeuLeu 396
 QY 317 CTATATGTTTATGATGACGATATATTATGATTTGGAGCC-----GTC 358
 DB 397 ValIeuValIyserleuLeuIytrYrYrValIeuIlyleuThralaLeuSerleu 416
 QY 359 ATGCTTGGGAGAGCTTCTTCTGAACTCTTCTTCTTCTTCTTCTTCTTCTTCTT 418
 DB 417 ValAlaserIlyserleuValPheAlaIlyleuValIeuLeuIly-----ArgTyrIle 434
 QY 419 GATGCGCACTACCTTGTGAGACTGCGTGGATGTTCTTCTTCTTCTTCTTCTT 478
 DB 435 AsnIytrThirleuaspLeuIlyleuIlyleuIlyleuIlyleuIlyleuIly 454
 QY 479 GATGCAATGCTCTGATTTGATTCGAAAGATCCGAGAGATTTTATGTCCTCAAGCTT 538
 DB 455 AspserheValValPhePheGlnIlyleuaspIlyleuArgIlyArgSerPhe 474
 QY 539 AAAAAATCTGTAGAAAAAGATATACCAAGCTTTTGGAGCCATTTTGAATCACTTG 598
 DB 475 ArgserIlyalValProargIlyTtpalArgalArgIlyThrIlyleuSerGlyAsnAla 494
 QY 599 ACTACGATTTGGCTCAGACTCTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 658
 DB 495 ValIytrheleuAlaIlyleuValIytrPheleuAlaIlyleuIlyGlnIlyGlyPhe 514
 QY 659 GCTTTCATGATTTTATGAAATTTCTCTCAATGTTTACGCTCTTTCATGACTTAA 718
 DB 515 Alaphethrleu-----GlyleuThrIlyleuaspLeuVal 528
 QY 719 TTTTCTTCATGCTGTG 736
 DB 529 ValPheleuValIytrPhe 534
 RESULT 23
 SECF BORBU STANDARD; PRT; 299 AA.
 AC 051597;
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Protein-export membrane protein secF.
 GN SECF OR BB0653.
 OS Borrelia burgdorferi (Lyme disease spirochete).
 OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
 OX NCBI_TaxID=139;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 35210 / B31;
 RX MEDLINE=98065943; PubMed=9403685;
 RA Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A.,
 RA Lathigra R., White O., Ketchum K.A., Dodson R., Hickey E.K., Gwinn M.,
 RA Dougherty B., Tomb J.-F., Fleischmann R.D., Richardson D.,
 RA Peterson J., Karpavavage A.R., Quackenbush J., Salzberg S., Hanson M.,
 RA van Vugt R., Palmer N., Adams M.D., Gocayne J.D., Weidman J.,
 RA Utterback T., Matthey L., McDonald L., Attiach P., Bowman C.,
 RA Garland S., Fujii C., Cotton M.D., Horst K., Roberts K., Hatch B.,
 RA Smith H.O., Venter J.C.;
 RT "Genomic sequence of a Lyme disease spirochete, Borrelia
 RT burgdorferi";
 RL Nature 390:580-586(1997).
 CC -1- FUNCTION: INVOLVED IN PROTEIN EXPORT (BY SIMILARITY).
 CC -1- SUBUNIT: PART OF THE PROKARYOTIC PROTEIN TRANSLLOCATION APPARATUS

Alignment Scores:

Pred. No.: 7.48e-15 Length: 442
 Score: 279.50 Matches: 91
 Percent Similarity: 44.33% Conservative: 67
 Best Local Similarity: 25.55% Mismatches: 125
 Query Match: 8.44% Indels: 73
 DB: 1 Gaps: 12

US-09-868-987-1 (1-1864) x SECF_MYCTU (1-442)

QY 794 CATGATTTCTTGAGA-----GSATGCAA 817
 DB 45 HisSerPheLeuSerArgLeuThrGlyThrGlyAlaPheGluValAlaGlyArg 64
 QY 818 AAACCTTGG--GCTGTTCTGGAAGTGTTCCTTTAGGTGGTCTCGCGGTT 874
 DB 65 ArgLeuTrpPheGlyValSerGlyAlaLeuAlaAlaLeuAlaLeuSerLeuValPhe 84
 QY 875 GGAGCCTGGAATTCGCTTTGGGAATGATTTAAAGAGGGTATGCTTACCTTAA 934
 DB 85 ArgGlyPhe-----ThPheGlyLeuAspPheGlyGlyThrThrValSerPhe--- 101
 QY 935 CCAAAAGACATGGCATGAGCATGTGCTCAATGCGTGGCAAAAGTTGATGATTA 994
 DB 102 -----ProArgGlySerThrGlnValAlaGlnValAlaGluAspValTyrTyrArgAlaLeu 119
 QY 995 CAGGAAGCTGGTCTTCTTCTAGACATTCGATTCACAACTTGGATCTTCAGAAA 1054
 DB 120 -----GlySerGluProGlnSerValValLeuAlaGlyAlaGlyAlaSerAlaThr 136
 QY 1055 ATCAAAATCTATTATGATTAAGATTAAGCTTAAAGCTTACTAAGACATACGACCTCTC 1114
 DB 137 ValGlnIleArgSerGluThrLeuThrSerAspGlnThrAlaGlyLeuArgAspAlaLeu 156
 QY 1115 CTAATAATTAAGATCATGATGAGCGGCTTATGTGGATTTGTGCAGAAAGCGCTTGA 1174
 DB 157 PheGluAlaPhe-----GlyProlys 163
 QY 1175 TTTCTCTACGGAACCTTAACGAAACGCAAAATTTGTGCAAGTAGACGCAAACTA 1234
 DB 164 GlyThrAspGlyGlnPrSerIleGlnAlaIleSerAspSerAlaValSerGluThrTrp 183
 QY 1235 TCGAAGAAATGCGTTATCAGGCCACATCGGCTTTTAAAGAGCTTTGGCATCATCTTG 1294
 DB 184 GlyGlyGlnIleThrLeuLeuValAlaLeuValAlaPheLeuValLeuValAla 203
 QY 1295 CTTATGATGAGTTGGCTTGAATGAGCAATATGCTTCAGGCGGTATGCGCTTAAT 1354
 DB 204 LeuTyrIleThrValAlaIleTyrGluArgTyrIleThrIleSerAlaIleThrAlaMetLeu 223
 QY 1355 CATGACCTTTGGCTACCTGTGACACTGTGTTATAGACATTTCTTTTGAAGAAAT 1414
 DB 224 PheAspLeuThrValThrAlaGlyValIleGlySerLeuValGlyPhe----- 238
 QY 1415 CAATAGATTGGCAAGCATGTCCTTAAATGACTGATTTGGGCTATTCATTAACAAT 1474
 DB 239 GluValThrProAlaThrValIleGlyLeuLeuThrIleLeuGlyPheSerLeuThrAsp 258
 QY 1475 ACTTGATCATTTTGAATGATGATGATGATGATGATGATGATGATGATGATGAT 1507
 DB 259 ThrValIleValPheAspLeuValGluGluLeuThrHisGlyPheGlnHisThrThrArg 278
 QY 1508 -----CGCCAAGCAAGCAAGCTGTTTACCCCTATGATGATTTTAAATGATGCC 1555
 DB 279 ArgThrPheAlaGluGlnAlaAsnLeu-----Ala 288
 QY 1556 CTTCAAAAGAGCTTCAGCGCAGGTAATGACACAGCTACA--ACTCTATCAGTTTG 1612
 DB 289 IleAsnGlnThrPheMetArgSerIleAsnThrSerLeuIleGlyValLeuProValLeu 308
 QY 1613 TTAAGCTTTTG-----TTTATAGCGGCTCTCTGCTTAAATTTGATTT 1660

DB 309 AlaLeuMetValAlaValAlaValTrpLeuLeuGlyValGlyThrLeuLysAspLeuAlaLeu 328
 QY 1661 ATTAGACCACTAGGATTTCTTCTAGCACTTATTCGCTCTTATATGACACACCTCTG 1720
 DB 329 ValGlnLeuIleGlyIleIleIleGlyThrTyrSerSerIlePhePheAlaTrpProLeu 348
 QY 1721 TTGTGTTT-----ATGTCGCGTAAGAAATAGCC 1750
 DB 349 LeuValThrLeuArgGluArgGluLeuValArgGlnHisThrArg 364
 RESULT 25
 SECD STRCO STANDARD; PRT; 570 AA.
 AC Q53955; Q9L293;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Protein-export membrane protein secd.
 GN SECD OR SCO1516 OR SCL2.06C.
 OS Streptomyces coelicolor.
 OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
 OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomycetes.
 NC NCBI_TaxID=1902;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3 (2).
 RA Lortieux A., Fraire P., Brans A., Dussart J.,
 RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3 (2) / M145;
 RX MEDLINE=21996410; PubMed=1200953;
 RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
 RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kleser H.,
 RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
 RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth M.,
 RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
 RA Rabinowitz E., Rajandream M.A., Rutherford K., Rutter S.,
 RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
 RA Warren T., Wietzorrek A., Woodward J., Barrrell B.G., Parkhill J.,
 RA Hopwood D.A.;
 RT "Complete genome sequence of the model actinomyete Streptomyces
 RT coelicolor A3(2)";
 RL Nature 417:141-147(2002).
 CC - FUNCTION: INVOLVED IN PROTEIN EXPORT.
 CC - SUBUNIT: PART OF THE PROKARYOTIC PROTEIN TRANSLLOCATION APPARATUS
 CC WHICH COMPRISE SECA, SECB, SECD, SECE, SECF, SECG AND SECH
 CC (BY SIMILARITY).
 CC - SUBCELLULAR LOCATION: Integral membrane protein. (By similarity).
 CC - SIMILARITY: BELONGS TO THE SECD/SECF FAMILY. SECD FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: X85969; CA5954.1; -
 DR EMBL: AL137778; CAB70918.1; -
 DR InterPro: IPR001036; Acrlvln_res.
 DR InterPro: IPR003335; Secd_Secf.
 DR Pfam: PF02355; Secd_Secf.1.
 DR PRINTS: PR00702; ACRIFLAVINRP.
 DR TIGRFAMs: TIGR00916; 2A0604801; 1.
 DR TIGRFAMs: TIGR01129; secd; 1.
 KW Protein transport; Translocation; Transmembrane; Membrane;
 KW Complete proteome.
 FT TRANSMEM 370 390 POTENTIAL.
 FT TRANSMEM 395 415 POTENTIAL.
 FT TRANSMEM 419 439 POTENTIAL.
 FT TRANSMEM 474 494 POTENTIAL.

FT TRANSMEM 498 518 POTENTIAL.
 FT CONFLICT 72 72 Y -> N (IN REF. 1).
 FT CONFLICT 191 191 S -> K (IN REF. 1).
 FT CONFLICT 550 551 PL -> SV (IN REF. 1).
 SQ SEQUENCE 570 AA; 59511 MW; 073ADE27F2B921E3 CRC64;

Alignment Scores:
 Pred. No.: 1.34e-14 Length: 570
 Score: 276.50 Matches: 68
 Percent Similarity: 54.31% Conservative: 58
 Best Local Similarity: 29.31% Mismatches: 91
 Query Match: 8.33% Indels: 15
 DB: 1 Gaps: 4

US-09-868-987-1 (1-1864) x SEC2_STRCO (1-570)

Qy 74 CQTATGCTGTAGTATGACGGTGTATATGTCAGCAGCCCTATTATAACGTCCTCATG 133
 Db 300 GlnPheAlaIleValLeuAspAsnGluValValSerAspProTyrValSerGlnAlaLeu 319
 Qy 134 AAA---AATCATGCCAGTGTCTCAGGAGAAATTTACCCACCGTGAAGTGGACAACTGCC 190
 Db 320 ThrGlyGlyAsnAlaGluIleSerGlySerPheAspGlnGluAlaGlnSerLeuAla 339
 Qy 191 TCAGATTTAAATCTGGAGCGATGCTTTTGTCCGAGGTTCTCAGTGAAGAGAGATC 250
 Db 340 AsnMetLeuSerTyrGlyAlaLeuProLeuThrPheLysGluAspSerValThrVal 359
 Qy 251 TCTCTGATCTGGGAAAAACAATGTACACAGGCAATATCTCAGCATGCTGTGGCTTG 310
 Db 360 ThrAlaAlaLeuGlyGlyGlnLeuLysAlaGlyLeuIleAlaGlyAlaIleGlyLeu 379
 Qy 311 GCAATGCTTATTGTTTGTAGCGCTATATATATATGATTTGGAGGCGTCTCGCGGA 370
 Db 380 AlaLeuValValLeuTyrLeuLeuPheTyrArgGlyLeuSerPheIleAlaValCys 399
 Qy 371 GCTGTTCTTGAATCTTTGTTGTTATCTGGGAGCTCTACAGTATTGGATGCG----- 424
 Db 400 SerLeuLeuValSerAlaGlyLeuThrTyrValIleMetAlaLeuLeuGlyProThrIle 419
 Qy 425 -----CCACTCACCTCTCAGGAGCTCGTGGATGTCTTGTCTATGGGATGCGCGTA 478
 Db 420 GlyPheAlaLeuAsnLeuProAlaValCysGlyAlaIleValAlaIleGlyIleThrAla 439
 Qy 479 GATGCAAAATGTTCTGTATTCGAAAGATCCGAGAGAAATTTTATTGCTCAAAAGTCTT 538
 Db 440 AspSerPheIleValTyrPheGluArgValArgAspGluIleArgGluGlyArgThrLeu 459
 Qy 539 AAAAAATCTGTAGAAAAGGATATACCAAGGCTTTTGGAGCCATTTTGTATCTAACTTG 598
 Db 460 ArgProAlaValGluArgAlaTyrProArgAlaArgThrIleLeuValSerAspPhe 479
 Qy 599 ACTACAGATTGGCTCTCAGCATCTTTTCTTCCTAGATACAGGCGCTTAAAGGTTT 658
 Db 480 ValSerPheLeuAlaAlaValLeuPheIleValThrValGlyLysValGlnGlyPhe 499
 Qy 659 GCTTTGACATTTG-----ATTTTAGGAATTTTCTCTTCA 691
 Db 500 AlaPheThrLeuGlyLeuThrThrLeuLeuAspValValValPheLeuPheThrLys 519
 Qy 692 ---ATGTTTACGGCTCTTTTCATGACTAAATTTTTC 724
 Db 520 ProLeuLeuThrLeuMetAlaArgArgLysPhePhe 531

RESULT 26
 SEC2_TREPA
 ID SEC2_TREPA STANDARD; PRT; 420 AA.
 AC O83426;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Protein-export membrane protein secf.
 GN SEC2 OR TP0411.

OS Treponema pallidum.
 OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.
 OX NCBI_TaxID=160;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Nichols;
 RX MEDLINE=98332770; PubMed=9665876;
 RA Fraser C.M., Norris S.J., Weinstock G.M., White O., Sutton G.G.,
 RA Dodson R., Winn M., Hickey E.K., Clayton R., Ketchum K.A.,
 RA Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,
 RA Khalak H., Richardson D., Howell J.K., Chidambaram M., Utterback T.,
 RA McDonald L., Artiach P., Bowman C., Cotton M.D., Fujii C., Garland S.,
 RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,
 RA Venter J.C.;
 RT "Complete genome sequence of Treponema pallidum, the syphilis
 RT spirochete";
 RL Science 281:375-388(1998).
 CC -1- FUNCTION: INVOLVED IN PROTEIN EXPORT.
 CC -1- SUBUNIT: PART OF THE PROKARYOTIC PROTEIN TRANSLLOCATION APPARATUS
 CC WHICH COMPRISE SECA, SECB, SECD, SECF, SECG AND SECH
 CC (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).
 CC -1- SIMILARITY: BELONGS TO THE SEC2/SECF FAMILY. SEC2 FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL; AE001219; AAC65399.1; -.
 CC TIGR; TP0411;
 DR InterPro; IPR003335; Sec2_SecF.
 DR Pfam; PF02335; Sec2_SecF; 1.
 DR TIGRFAMS; TIGR00916; 2A0604801; 1.
 DR TIGRFAMS; TIGR00966; 3A0501807; 1.
 KW Protein transport; Translocation; Transmembrane; Membrane;
 KW Complete proteome.
 KW TRANSMEM 7 27 POTENTIAL.
 FT TRANSMEM 250 270 POTENTIAL.
 FT TRANSMEM 276 296 POTENTIAL.
 FT TRANSMEM 297 317 POTENTIAL.
 FT TRANSMEM 358 378 POTENTIAL.
 FT TRANSMEM 388 408 POTENTIAL.
 SQ SEQUENCE 420 AA; 45731 MW; D7FC706AF596F6F1 CRC64;

Alignment Scores:
 Pred. No.: 4.76e-14 Length: 420
 Score: 269.50 Matches: 67
 Percent Similarity: 60.59% Conservative: 36
 Best Local Similarity: 39.41% Mismatches: 50
 Query Match: 8.14% Indels: 17
 DB: 1 Gaps: 6

US-09-868-987-1 (1-1864) x SEC2_TREPA (1-420)

Qy 1220 GTAGAGCAAACTATCGAAGAAATCGTGTATCAGGCGACCATCGGCTTTTAGGAGCT 1279
 Db 243 ValGlySerArgPheSerHisLeuLeuValArgGlnAlaLeuLeuValLeuGlyAla 262
 Qy 1280 TTGGCAATCATCTTGTCTATGTCAGTTTGGCTTTGAATGGCAATATGCTTTCAGTGCC 1339
 Db 263 LeuValLeuIlePheLeuTyrValAlaLeuArgPheArgTrpPheAlaLeuGlyAla 282
 Qy 1340 GTATGCGCTTTAAATTCATGACCTTTTGGCTGTGTCAGTCTTGTATAGCACATTC 1399
 Db 283 IleValAlaLeuValHisAsp-----AlaCysIleMetValSerPheMetValTrpPhe 300
 Qy 1400 TTTTGTGAAGAAATTCAAATAGATTTCAGAGCCATTGGTGTCTTAACTGCTGTTATGGG 1459
 Db 301 GlyLeu-----GluPheAsnSerAlaSerIleAlaAlaIleLeuThrIleIleGly 317

Pred. No.:	4,688-13	Length:	315
Score:	257.00	Matches:	84
Percent Similarity:	43.14%	Conservative:	67
Best Local Similarity:	24.00%	Mismatch:	141
Query Match:	7.76%	Indels:	58
DB:	1	Gaps:	9

US-09-868-987-1 (1-1864) x SECF_SYNY3 (1-315)

OY	788	ATAAAGCATGATTTCTTGAGAGATGCAGAAAACTTGGCGCTGTTCGAAAGTGTTTT	847
Db	1	MethylLeuaspLeuPheNylSTRPGlulysProAlatrpIleValSerSerLeuVal	20
OY	848	CTTTTAGGTTCCGTTCGCTCCGGGTTTGAGCCTCGGAATTCGTTTTG-----	895
Db	21	LeuileSerIlePheAlaMetAlaIle--SetrIpIagInPheGlInAlaPropheaArg	39
OY	896	---GGAATGATTTTAAGAAGGGGTATGCTTACCTTAAT-----	934
Db	40	ProdllyLeuaspPheValIGlyIthrArgLeuGlInleuGlInleuGlucySalasSer	59
OY	935	-----CCAAAAGCATGCGCATCAGCGATGTGTCTAAATGCGTGCAGAAAGTTTG	985
Db	60	AsnaIncysProAlaIalaleaspValAlaGlulValGlInaspIleLeuGlIglyAl---	78
OY	986	CATTAACACTACAGAAAGCTGGCTTTCTTCTTCAAGACTCCGATTCATAAACATTGGATCT	1044
Db	79	-----GlyLeuGlIynSerSerValGlInAl-----	87
OY	1046	TCAAAAAGATCAAATATCTATTTAGTGAATAAAGCTTAAAGCTATCTAAGCAGATACGA	1107
Db	88	-----IlglubspIyrThrLeuSerIleArg	96
OY	1106	GCCCTCTCCTAAATATMAAGCATCAGCTGCGTTATVTGGAGATGTGTCAAGAAC	1161
Db	97	GlInGIInThrLeuasp-----ValGIuGIIn	104
OY	1166	AGCGCTTGATTTCTTACGGAAACTTAAACGAAACGCCAAATTT-----	1211
Db	105	ArgGlulAlaValGIInlysAlaLeuamngluGIyIleGIlySpheaspProGIuthrIle	124
OY	1211	---TGRCAAMAGGAGACAGCAACAATCGAAGAAATACGTTATCAGCGCACATCGGG	126
Db	125	GlInIleaspThrValGIyProthrValGIyLysAlaLeuPheArgSerGIyValIleuAla	144
OY	1268	CTTTTAGAGCTTTGGCAATCATCTTGCTATGTAGTTCGGCTTGAATGSGCAAT	1327
Db	145	LeuValIleSerLeuLeuGIyIleIleIleIyLeuThrIleArgPheInleuaspIyr	164
OY	1328	GCITTCAGTCCCGATGCGCTTAAATTCATGACCTTTGGCTACCTGTGCACTGTGTT	1387
Db	165	AlaValPheAlaIleIleAlaIleLeuIyryAspAlaLeuIleThrmecGIyAla--Phe	183
OY	1388	ATAGACATTTCTTTTGAAGAAATTCAAATRGATTTGCAAGCCATTGSGCTTAAATG	144
Db	184	AlaIlePheGIyLeuValGIyGIyValGIuValAspSerLeuPheLeuValAlaIleuLeu	203
OY	1448	ACTGATTTGGGATTCATTAACATCACTTGTGATCTATTTGGATGGTATGCGTGAAGAT	1507
Db	204	ThrIleIleGIyPheSerValInaspThrValIleIyryAspArgValArgGIuthr	223
OY	1508	CGCCACGCAACCTGTTTACCCTATGCATGTTTAAATGATGACCCTTCMAAAGACG	1567
Db	224	LeuGIulnArgHisSerAspTrpAspIleAmnHISvalValAspAspAlaValAmngInThr	243
OY	1568	TTACAGCCGCAAGGTAATGACACAGCTACACACTGTATCAGTTTGTAAATGCTTTGTTT	1622
Db	244	LeuThrArgSerIleasnThrSerLeuThrThrSerLeuProLeuValAlaIlePheLeu	263
OY	1628	ATAAGCGGCTCCCTGCTTAAATTTTGGATTTATATGACATATGGAGATCTTCTAGGA	1687
Db	264	PheGIyGIyAspSerLeuIynSphPheAlaLeuAlaIleIleGIyPheAlaSerGIy	283

Qy 1688 ACTTATCGTCTCTTATATATGACACCTCTGTGTGTTTATGTCGCGTAAGAAAT 1747
 Db 284 ValTyrSerSerIlePheMetAlaThrThrLeuTrpAlaTrpTrp---ArgLysTrpArg 302
 Qy 1748 CGCTCAAAATAGTACCGTTAACTTAATC 1777
 Db 303 SerProLysAsnProProArgGluMetVal 312

RESULT 28 SECF_STRCO

ID SECF_STRCO STANDARD; PRT; 373 AA.
 AC Q53956;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-JUN-1997 (Rel. 35, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Protein-export membrane protein secf.
 GN SECF OR SCO1515 OR SCL2.05C.
 OS Streptomyces coelicolor.
 OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
 OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
 OX NCBI_TaxID=1902;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA Loriaux A., Frere P., Brans A., Dusart J.;
 RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2) / M145;
 RX MEDLINE=21996410; PubMed=12000953;
 RA Bentley S.D., Chater K.F., Corden-Tarraga A.-M., Challis G.L.,
 RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
 RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
 RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
 RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neil S.,
 RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
 RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
 RA Warren T., Wietzorrek A., Woodward J., Barrrell B.G., Parkhill J.,
 RA Hopwood D.A.;
 RA "Complete genome sequence of the model actinomycete Streptomyces
 coelicolor A3(2).";
 RL Nature 417:141-147(2002).
 CC -!- FUNCTION: INVOLVED IN PROTEIN EXPORT.
 CC -!- SUBUNIT: PART OF THE PROKARYOTIC PROTEIN TRANSLOCATION APPARATUS
 CC WHICH COMPRISE SECA, SECB, SECD, SECE, SECF, SECG AND SECY
 CC (BY SIMILARITY).
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).
 CC -!- SIMILARITY: BELONGS TO THE SECD/SECF FAMILY. SECF FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL; X85969; CAA59955.1; -
 CC EMBL; ALJ37778; CAB70917.1; -
 CC InterPro; IPR003335; Secd_Secf.
 CC Pfam; PF02355; Secd_Secf; 1.
 CC TIGRFAMs; TIGR00916; 2A0604s01; 1.
 CC TIGRFAMs; TIGR00966; 3A0501s07; 1.
 CC Protein transport; Translocation; Transmembrane; Membrane;
 KW Complete proteins.
 KW TRANSMEM 26 46 POTENTIAL.
 FT TRANSMEM 142 162 POTENTIAL.
 FT TRANSMEM 166 186 POTENTIAL.
 FT TRANSMEM 193 213 POTENTIAL.
 FT TRANSMEM 251 271 POTENTIAL.
 FT TRANSMEM 280 300 POTENTIAL.
 FT TRANSMEM 373 AA; 39710 MW; 28C55B23C62F82ED CRC64;
 SQ SEQUENCE

Alignment Scores:
 Pred. No.: 5,75e-13 Length: 373
 Score: 256.00 Matches: 86
 Percent Similarity: 43.73% Conservative: 64
 Best Local Similarity: 25.07% Mismatches: 125
 Query Match: 7.73% Indels: 68
 DB: 1 Gaps: 13

US-09-868-987-1 (1-1864) x SECF_STRCO (1-373)

Qy 788 ATAAAGCATGATTTCTTTGAGAGGATGCAAAAACCTTTGGGCTGTTTCTGGAAGTGTTTT 847
 Db 16 ValGlyTyrAspPheValLysAsn---ArgLysIleTrp---TyrGlyIleSerIleLeu 33
 Qy 848 CTTTAGGTTCGGTGTCTCTCGGGTTTGGAGCCCTGGAAATCCGTTTGGGAATGGATTTT 907
 Db 34 IleThrIleThrAlaIleValGlyLeuAlaValArgGlyLeuHisMetGlyIleGluPhe 53
 Qy 908 AAAGGAGGTATGCTTTTACCTTTAATCCAAAAGAGCATGCGATCAGC----- 955
 Db 54 GluGlyGlyAlaValPheThr---ThrProLysAsnMetSerAlaSerValAlaGlnThr 72
 Qy 956 -----GATGTTGCTCAAAATCGGTGGCAAACTTGTGCATAACTACACGAAGCT 1003
 Db 73 GluThrTrpAlaGluAlaSerGlyHisAspAlaIleValGlnLysLeuGlyAspGly 92
 Qy 1004 GGTCTTTCTTAGAGACTTCCGTATTCAACACATTTTCGATCT-----TCAGAA 1051
 Db 93 SerLeu-----ArgIleGlnIleAlaGlyThrAspThrGlnGlnSerAsp 107
 Qy 1052 AAGATCAAAATCTATTATTAGTATAAAGCTTTAAGCTATCTACTAAGCAGATACGAGCTCT 1111
 Db 108 GlnIleLysGluAspLeuSerLysAsnLeuAspValSerAlaGluLysIleAsnAlaAsp 127
 Qy 1112 CTCTAAATTAACGATCATGAGCTGCGGTATTGTGCGGATTGTGTGCAGAAACAGCCT 1171
 Db 128 LeuValGly-----ProSerTrp----- 133
 Qy 1172 AGATTCTCTACGGAAACTCTAAACGAAAGCGCAAAATTTTGGTCAAAAGGTAAAGCAGCAA 1231
 Db 134 -----GlyAspGlnIleAlaAsn---LysAlaTrpGln----- 143
 Qy 1232 CTATCGAAGAAATATGCTTATACGGCGACCATCGGGCTTTTAGGAGCTTTGGCAATCATC 1291
 Db 144 -----GlyLeuGlyIlePheMetValLeuVal 152
 Qy 1292 TTGCTCTATGTAGTTTGGCTTTGAATGCGCAATATCTTTAGTGGCTATGCGCTTTA 1351
 Db 153 ValIleTyrLeuAlaIleAlaPheGluTrpArgMetAlaLeuAlaPheValAlaLeu 172
 Qy 1352 ATTCATGACCTTTTGGCTACTGTGCGAGCTGTGTTTATAGCACATTTCTTTTGAAGAAA 1411
 Db 173 IleHisAspIleThrIleThrValGlyIleTyrAlaLeuValGlyPhe----- 188
 Qy 1412 ATTCAAATAGATTTCGAAGCCATTTGCTTTAATGACTGTATTGGGGTATTCATTAAAC 1471
 Db 189 ---GluValThrProGlyThrValIleGlyLeuLeuThrIleLeuGlyTyrSerLeuTyr 207
 Qy 1472 AATACTTTGATCATTTTGTGATCGTATTCGTGAA-----GATCCCAAGCAACCTG 1522
 Db 208 AspThrValValValPheAspSerLeuLysGluGlnThrArgAspIleThrLysGlnThr 227
 Qy 1523 TTTACCCCTATGCATGTTTGTAGTTAATGATGCGCTTCAAAAGACGTTTCAGCGCAGCGTA 1582
 Db 228 ArgTrpThrTyrAlaGluIleAlaAsnArgSerIleAsnSerThrLeuValArgSerIle 247
 Qy 1583 ATGACAACAGCTACAACCTCATCATGTTTGTGTTAATGCTTTTGTATAGGCGGC----- 1636
 Db 248 AsnThrThrValValAlaLeuLeuProValAlaGlyLeuLeuPheIleGlyGlyVal 267
 Qy 1637 -----TCCTCTGCTCTTAATTTTGCATTTTATATGACCATAGGATTCCTCTAGGA 1687
 Db 268 LeuGlyAlaGlyMetLeuAsnAspIleSerLeuSerLeuPhePheValGlyLeuAlaGly 287

```

Oy 1688 ACTTATACGCTCTCTTATATGACACACCGCTGTGTATATGCTCCGTAAGAANAAT 1747
      |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 288 AATATrSeSeStIleheHeHeaIaHrPrOLEuValAlAspLeuYsgUlaGIuPPro 307
      ::|||
Oy 1748 CGCTCAAAA 1756
      ::|||
Db 308 GImeLtyS 310
      ::|||

RESULT 29
SECF_MYCLE STANDARD: PRT; 471 AA.
ID SECF_MYCLE STANDARD: PRT; 471 AA.
AC P38386; 069486;
DT 01-OCT-1994 (Rel. 30, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Protein-export membrane protein secf.
OS SECF OR MU0488 OR MCB1259..06 OR B1177_G3_239.
OS Mycobacterium leprae.
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteriidae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;
RN [1]
RP SEQUENCE FROM N.A.
RA Smith D.R., Robison K.;
RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=TN;
RX MEDLINE=21128732; PubMed=11234002;
RA Cole S.T., Eglmeier K., Parkhill J., James K.D., Thomson N.R.,
RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
RA Holroyd S., Hornby T., Jagsels K., Lacroix C., Maclean J., Moule S.,
RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
RA Rutter S., Seeger K., Simon S., Simmonds M., Skellon J., Squares R.,
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA Barrell B.G.;
RT "Massive gene decay in the leprosy bacillus.";
RL Nature 409:1007-1011(2001).
CC - FUNCTION: INVOLVED IN PROTEIN EXPORT.
CC - SUBUNIT: PART OF THE PROKARYOTIC PROTEIN TRANSLOCATION APPARATUS
CC WHICH COMPRISE SECA, SECB, SECD, SECE, SECF, SECG AND SECY
CC (BY SIMILARITY).
CC - SUBCELLULAR LOCATION: Integral membrane protein (by similarity).
CC - SIMILARITY: BELONGS TO THE SECD/SECF FAMILY. SECF FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation.
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL, U00011; AAA17100.1; ALT FRAME.
DR EMBL, AL023591; CA19081.1; --
DR EMBL, AL583918; CAC29996.1; --
DR Leptoma; MU0488;
DR InterPro; IPR003335; SeCD_SecF.
DR Pfam; PF02355; SeCD_SecF; 1.
DR TIGRFAMs; TIGR00916; 2A0604S01; 1.
DR TIGRFAMs; TIGR00966; 3a0501a07; 1.
KW Protein transport; Translocation; Transmembrane; Membrane;
KW Complete proteome.
KW TRANSMEM 88 108 POTENTIAL.
KW TRANSMEM 116 136 POTENTIAL.
KW TRANSMEM 211 231 POTENTIAL.
KW TRANSMEM 267 287 POTENTIAL.
KW TRANSMEM 325 345 POTENTIAL.
KW TRANSMEM 355 375 POTENTIAL.
KW SEQUENCE 471 AA; 5018 MW; 1ADB6D28297A31217 CRC64;

```

Alignment Scores:			
Pred. No.:	1.24e-12	Length:	471
Score:	252.00	Matches:	81
Percent Similarity:	44.55%	Conservative:	66
Best Local Similarity:	24.55%	Mismatches:	125
Query Match:	7.61%	Indels:	58
DB:	1	Gaps:	10
US-09-868-987-1 (1-1864) x SECF_MYCLE (1-471)			
QY 809	GGATGCAAAAACCTTTGG---GCTGTTTTCGA-----AGCTTTTCTTTAGCT		856
Db 79	GIYARgARgARgLeuRTPYrSerValGIgLYValIIleValAlaValAlaValLeuSer		98
QY 857	TGCGTCTCTCGGgTTTGGAGCGCTGCAATTCGGTTTGGGAATGATTTTAAAGGAGG		916
Db 99	IIleIIeValArgLYPhe-----ThrPheGIIYIIaSPhePheGLYGLY		113
QY 917	TATGCTTTTACCTTTATTCGAAAGACATGCGATCAGCGATGTTGTCGAAATGCGTGGC		976
Db 114	ThrThrValSerMet---ProValSerProGIgLYValGIgLYThrGIYAlaIIleGIUVal		132
QY 977	AAAGTTGCAATTAACACAGAGAGCTGGCTTTCTTCAGAGACTTCGCTATTCGAAACA		1036
Db 133	AlaGlnValAlaAspValPheYrSuThrIleuGIYSerASPProGIUSeValValVal		152
QY 1037	TTTGATCTTCAGAAAAGATCAAAATCTATTTTAGATTAAGCTTTAGC-----TAT		1090
Db 153	ValGIYsendlYAlaSerAlaThrValaGIleSerSerLYgThrLeuSerASPaspGln		172
QY 1091	ACTAAGCAGATACGAGCGCTCTCTCTTAAATTAACGATCATGAGCTGGCGTTATGTGGC		1150
Db 173	ThrSerLYsLeuArgSnaIIaLeuPheASPAlaPhe-----		184
QY 1151	ATTTGTTGCAAAACAGCGCTGATTTCTCTACGGAACCTCAACGAAACGCAAAATTT		1210
Db 185	-----GIYProLYsGIYAlaAspAlaLYsProSerLYsGlnAlaIIleSer		199
QY 1211	TGATCAAGGTAAAGCAGCAAACTATCGAAGAAATGCGTTATCAGGCAACATCGGCGTT		1270
Db 200	AspAlaAlaValaSerGIuThrTPeGIYGIYglnIIeThrLYsValValaIIleAlaLeu		219
QY 1271	TTAGAGCTTTGGCAATCATCTTGCTCTATGTGAGTTGGCTTGCGTTGAATGCAATATGCT		1330
Db 220	ValAlaPheLeuValLeuValGIYLeuTYrIIeThrValaArgTYGIaArgTYIeMetAla		239
QY 1331	TTTACGTGCGTAAACGGCTTTAATCATAGACCTTTGGGCTACTCGTGCAGCTGTGTTATA		1390
Db 240	IIleSerAlaLeuThrThrMetCYsPheASPLeuThrValThAlaGIYValTYrSerLeu		259
QY 1391	GCACATTTCTTTTGAAGAAATTCAAATAGATTTTGCAAGCCATGGTGCCTTTATGACT		1450
Db 260	ValGIYIYhe-----GIUValThrProAlaThrValIIleGIYLeuLeuThr		274
QY 1451	GTAATGGGGATTCATTAACAATCACTTTGATCATTTTGTATCGATTTCGTGAAGAT---		1507
Db 275	IIleleuGIYpHeSerLeuTYrASPThrValIIleValaPheASPLYsValGIUGIUsnThr		294
QY 1508	-----CGCAAGCAAGCACTGTTAACCCCT		1531
Db 295	HisGIYpHeGlnHisThrThrArgArgThrPheAlaGIUGlnAlaAsnLeu-----		311
QY 1532	ATGCAATCTTTAGTTATGATGCGCTTCAAAAGACCTTCAGCCGACGCGTATGACACA		1591
Db 312	-----AlaIIeAsnGIuThrPheMetArgSerIIeAsnThrSer		324
QY 1592	GCTTCAACCTCATACGATTTGTTAATGCTTTTG-----TTTATAGCGCGC		1636
Db 325	LeuIIleSerValaLeuProValaLeuAlaLeuMetValaValaValaTriPLeuLeuGIYVal		344
QY 1637	TTCCTGCTGCTTATATTTTGATATTATTAAGACATAGAGGATCTCTTGAACCTTATCG		1696

Db	185	AlaAspGlyLysProHisProGlnGlnValIleThrValGlyAlaTyrProProThrLys	204
QY	170	CGTGAAGTGGAGCAA-----CTCGCTCAGATTTAAATCTCGAGCGATGCTCTTT	230
Db	205	GlulGlulIleAspGluAlaMetAlaIleTyrSerAlaLeuLysSerGlyAlaLeuProVal	224
QY	221	GTTCCCGAGGTCTCAGTGAAGAGACGATCTCTTCTGATCTTGGGAAAAACAATGTACA	280
Db	225	LysLeuAspIleGluTyrIleSerThrIleSerProGluPheGlyLysGluPheLeuLys	244
QY	281	CAAGGCATATCTCAGCATGCTGCTGGCGCAATGCTATTGTTTGTATGACCGGTATAT	340
Db	245	GlyThrAlaIleAlaLeuLeuLeuAlaPheIleAlaValGlyIleIleValSerIleArg	264
QY	341	TATAGATTGAGGCGTCATCGCTCGGAGCTGTT-----CTTCTGAATCTTTTGCTT	394
Db	265	TyrLysGlnProLysIleAlaIleProIleLeuIleThrCysIleSerGluValIleIle	284
QY	395	ATCTGGGCAGCTCTACAGTATTTGGATGGCCACTCACCTTGTTCAGGACTCGCTGGGATT	454
Db	285	IleLeuGlyPheAlaSerLeuIleAspTyrLysLeuAspLeuProSerIleAlaGlyIle	304
QY	455	GTTCTTGCTATGGGATGGCCGTAGATGCMAATGTTCTTGTTATTCGAAGAATCCGAGAG	514
Db	305	IleAlaAlaValGlyThrGlyValAspAsnGlnIleValIleThrAsp-----	320
QY	515	GAATTTTATTGCTCAAGTCTTAAAAAATCTGTAGAAAA-----GGATATACC	565
Db	321	-----GluAlaLeuLysArgGlyAlaGlyLysIleArgAlaSerIleLys	335
QY	566	AAGGCTTTTGGAGCCATTTTGGATCTTAACCTGACTACAGTATTGGCCTCAGCACTCTT	625
Db	336	ArgAlaPhePheIleIlePheAlaSerAlaAlaThrSerIleAlaAlaMetLeuProLeu	355
QY	626	TTCTTCTAGATACAGGCTTATAAGGGTTTGTCTTGACATTTGATTTTAGGAATTTTC	685
Db	356	PheValLeuGlyValGlyMetLeuLysGlyPheAlaIleThrThrIleAlaGlyValLeu	375
QY	686	TCITCAATGTTT 697	
Db	376	IleGlyIlePhe 379	
RESULT 31			
Y849 METHTH			
ID Y849 METHTH	STANDARD;	PRT; 403 AA.	
AC Q26937;			
DT 30-MAY-2000 (Rel. 39, Created)			
DT 30-MAY-2000 (Rel. 39, Last sequence update)			
DE 16-OCT-2001 (Rel. 40, Last annotation update)			
DE Hypothetical protein MTH849.			
GN MTH849.			
OS Methanobacterium thermoautotrophicum.			
OC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;			
OC Methanobacteriaceae; Methanothermobacter.			
NCBI_TaxID=187420;			
[1] SEQUENCE FROM N.A.			
RP STRAIN=Delta H;			
RC MEDLINE=98037514; PubMed=9371463;			
RX Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., Dubois J. J.,			
RA Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,			
RA Harrison D., Hoang L., Keagle P., Lumm W., Pothier B., Qiu D.,			
RA Spadafora R., Voicare R., Wang Y., Wierzbowski J., Gibson R.,			
RA Jiwni N., Caruso A., Bush D., Sater H., Patwell D., Prabhakar S.,			
RA McDougall S., Shimer G., Goyal A., Pietrowski S., Church G.M.,			
RA Daniels C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.;			
RT "Complete genome sequence of Methanobacterium thermoautotrophicum			
RT deltaH: functional analysis and comparative genomics.";			
RT J. Bacteriol. 179:7135-7155(1997).			
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).			
CC -!- SIMILARITY: BELONGS TO THE SEC23/SEC24 FAMILY.			
CC			

QY 119 TTAACGCTCCATTTGAATAATCATCGCAGTGTCTCTCAGGAAATTTACCCACCGTGAAGTG 178
Db 111 LeuAsnValPheAsnLysGluAsnTyrGlnPheSerSerThrThrLysLeuVal 30
QY 179 AGCAAACTCGCCTCAGATTTTAAATCTGGAGCGATGCTTTTGTTCGAGGTTCTCAGT 238
Db 31 SerAsnSerThrValAspAlaAspAsnGlyAlaSerAspPhe-----GluAlaGlyGln 48
QY 239 GAAGAGACGATCTCTCTGATCTTTGGGAAACAAATGTCACACAGGCATTATCTCAGCA 298
Db 49 GlnPheAlaThrGluLeuAspGlnGlyGlyGlnLeu-----GlyLeuSer--- 65
QY 299 TGCTGTGCTGGCAATGCTTATTTGTTGAGCGATGATATATATAGATTGGAGGCGTC 358
Db 66 CysIleGlyLeuIleCysAsnArgMetLeuGlyThr-----GlyVal 79
QY 359 ATCGCTTCGGGAGCTGTTCTCTCAATCTT-----TTGCTTATC 397
Db 80 PheAlaValSerSerThrIleTyrThrLeuCysGlySerValGlyLeuAlaLeuIleMet 99
QY 398 TGGCAGCTCTACAGTATTTGGATGCGCCACTCACCTTGTGAGGCTCGCTGGGATTGTT 457
Db 100 TrpAla-----ValGlyAlaIleIleAlaIleSerGlyLeuTyr---ValTyr 114
QY 458 CTGCTATGGGATGGCGGTAGATGCAAAATGTTCTGTATTTCGAAAGATCCGAGAGGAA 517
Db 115 MetGluPheGlyThrAlaIlePro----- 122
QY 518 TTTTATTGCTCAAGCTTTAAATAATCTGTAGAAAAGATATACCAAGCTTTTGA 577
Db 123 -----LysAsnGlyGlyGluLysAsnTyrLeuGlu----- 132
QY 578 GCCATTTT-----GATTCTTAACCTGACTACAGTATTGGCTCAGCAGCTTCTTTCTTC 631
Db 133 AlaIlePheArgLysProLysPhePheIleThrCysMetTyrAlaAlaTyrIlePhePhe 152
QY 632 CTAGATCAGGCTTATAAGGTTTGTCTTGTGATGATTGATTAGGAATTTCTCTTCA 691
Db 153 Leu-----GlyTrpAlaAlaGlyAsnSerIleAsnThrAlaIleMetPheLeuThrAla 170
QY 692 ATGTTTACGGCTCTTTTCATGACTAAATTT----- 721
Db 171 AlaAspThrGlu-----ValThrLysTrpAsnGlnArgGlyIleGlyValAlaVal 188
QY 722 TTCCTTCATGCTGTGGATGAATAAGACCAACATACACAGTTGCATATGATGAATAAGTTC 781
Db 189 PhePheAlaPheLeuIleAsnSer-----LeuAsnValLys 200
QY 782 GTGGGATAAAGCATGATTTCTTGAGAGGATGCAAAAACCTTTGGCTGTTTCTGGAAGT 841
Db 201 IleGlyLeu-----TyrLeuGlnAsn----- 207
QY 842 GTTTTCTTTTAGCTTGGTGTCTCGGTTTGGAGCTGGAATTCGTTTGGGAATG 901
Db 208 -----IleLeuGlyIlePheLysIleGlyIleValLeuPheIleSerIleThrGlyTrp 225
QY 902 GATTTTAAAGAGGCTATGCTTTACCTTTTAACTCAAAAGACATGCGATCAGCGATGTT 961
Db 226 ValAlaLeuGlyGly-----GlyLeuLysAsp--- 234
QY 962 GCTCAATGCGTGCAAGTTGTGCTAAACTACAGGAAGCTGCTTCTTCTTAGAGAC 1021
Db 235 -----GlyTyrGlnSerIleAsn 240
QY 1022 TTCGCTATTCAACATTTGGATCTTTCAGAAAAGATCAAAATCTATTTTAGTGATAAGCT 1081
Db 241 PheArg---AsnAlaPheGluGlyThrGluThrAlaThrAlaTyrGlyIleValAsnAla 259
QY 1082 TTAGCTATATAAGCAGATACGAGCTCTCTCTCTAAATTAACGATCATGAGCTGGCGT 1141
Db 260 Leu---TyrSer-----ValIleTrpSer 266

QY 1142 TATTGTGGGATTGTTGTCAAGAAACAGCGCCTAGATTCTCTCAGGAAACTCTAAACGAAAC 1201
Db 267 PheValGly----- 269
QY 1202 GCAAAATTTTGGTCAAAAGGTAAAGCAGCAAACTACGAGAAATAATCGCTTATCAGCGCGACC 1261
Db 270 -----TyrSerAsnValAsnTyrAlaLeuGlyGluValLysAsnProValArgThr 286
QY 1262 ATCGGCTTTTAGGAGCT-----TTGCAATCATCTTGTCTCTATGTGAGT 1306
Db 287 LeuLysIleAlaGlyProThrSerMetValPheLeuAlaIleIleTyrIlePheValAsn 306
QY 1307 TTGGCTTTGAATGGCAATATGCTTTCAGTGCCGATGCGCTTTTAATTCATGACCTTTTG 1366
Db 307 Ile-----AlaTyrPheAlaVal-----ValProLysAspLysLeu 318
QY 1367 GCTACCTGTGCGAGTCTTGTATTATAGCATTCTTCTTTTGAAGAAATTCAAATAGATTG 1426
Db 319 IleSerSerLysLeuIleLeuAlaAlaAspPhe-----AspIleValPheGlyGly 336
QY 1427 CAAGCCATTGCTGCTTAATGACTGTATTGGGGTATTCA---TTAAACAATACTTTG--- 1480
Db 337 GlnAlaLysArgAlaAlaAlaLeuValGlyLeuSerAlaLeuGlyAsnValLeuSer 356
QY 1481 ATCATTTTTCATCGTATTCGTGAAGATCGCAA----- 1513
Db 357 ValIlePheSerGlnGlyArgIleIleGlnGlnLeuGlyArgGluGlyValLeuProPhe 376
QY 1514 GCGNACTGTTTACCCCTATGCTATGCTTTTAACTGATGCCCTTCAAAAGACGTTTCAGC 1573
Db 377 SerAsnPhePheAlaSerSerLysProPhe---AsnSerProMetValGlyLeuPheGln 395
QY 1574 CGCACGGTAAATGACACAGCTACAACTCTATCA-----GTTTGTG 1612
Db 396 HisPheIleValCysThrValThrIleLeuAlaProProGlyAspAlaTyrLeuLeu 415
QY 1613 TTAATGCTTTTGTATTAGCGGCTCTCTGCTTTTAACTTGTGATTTTATGACCAT 1672
Db 416 ValGlnAsnLeuIleSerTyrProMetAsnIleAlaAsnPheAla-----IleSerAla 433
QY 1673 GGGATTCCTT----- 1681
Db 434 GlyLeuLeuTrpIleTyrTrpGlnArgArgGlnGlyLysIleGluTrpAsnProPhe 453
QY 1682 -----CTAGGAACCTTATCGTCTCTTTAT-----ATTGCA 1711
Db 454 LysAlaGlyValPheValThrGlyPhePheThrLeuSerAsnLeuTyrLeuIleAla 473
QY 1712 CCA 1714
Db 474 Pro 474
RESULT 33
ACRF_ECOLI
ID ACRF_ECOLI STANDARD; PRT; 1034 AA.
AC P24181;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 15-OCT-2001 (Rel. 40, Last annotation update)
DE Acriflavine resistance protein F (Protein envd).
GN ACRF OR ENVd OR B3266.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RA Xu J., Nilles M.L., Bertrand K.P.;
RL Submitted (XXX-1992) to the EMBL/GenBank/DBJ databases.
RN [2]
RP PRELIMINARY SEQUENCE FROM N.A.
RC STRAIN=K12;

RX MEDLINE=92079901; PubMed=1720861;
 RA Klein J.R., Heinrich B., Piapp R.;
 RT "Molecular analysis and nucleotide sequence of the envC operon of
 RT Escherichia coli.";
 RL Mol. Gen. Genet. 230:240-240(1991).
 RN [3]
 RP REVISIONS.
 RX MEDLINE=94012493; PubMed=8407802;
 RA Ma D., Cook D.N., Alberti M., Pon N.G., Nikaide H., Hearst J.E.;
 RT "Molecular cloning and characterization of acra and acrf genes of
 RT Escherichia coli.";
 RL J. Bacteriol. 175:6299-6313(1993).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blatner F.R., Plunkett G. II, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.P.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12.";
 RL Science 277:1453-1474(1997).
 CC -1- FUNCTION: INVOLVED IN CELL ENVELOPE FORMATION. IS PRODUCED IN
 CC EXTREMELY LOW AMOUNTS.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane.
 CC -1- SIMILARITY: BELONGS TO THE ACRB/ACRD/ACRF FAMILY.
 CC -1- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO
 CC FRAMESHIFTS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M96848; AAA02932.1; -;
 DR EMBL; X57948; CAA41017.1; -;
 DR EMBL; U18997; AAB58070.1; -;
 DR EMBL; AE000405; AAC76298.1; -;
 DR PIR; S18537; S18537;
 DR Ecogene; EG10267; acrf;
 DR InterPro; IPR001036; ACRflavin_res.
 DR InterPro; IPR004764; HAE1.
 DR Pfam; PF00873; ACR_tran; 1.
 DR PRINTS; PR00702; ACRIFLAVINRP.
 DR TIGRFAMs; TIGR00915; 2k0602; 1.
 KW Cell division; Transmembrane; Inner membrane; Transport;
 KW Complete proteome.
 FT DOMAIN 1 9
 FT TRANSMEM 10 28
 FT DOMAIN 29 339
 FT TRANSMEM 340 359
 FT DOMAIN 360 365
 FT TRANSMEM 366 385
 FT DOMAIN 386 391
 FT TRANSMEM 392 413
 FT DOMAIN 414 441
 FT TRANSMEM 442 460
 FT DOMAIN 461 473
 FT TRANSMEM 474 496
 FT DOMAIN 497 537
 FT TRANSMEM 538 556
 FT DOMAIN 557 871
 FT TRANSMEM 872 891
 FT DOMAIN 892 897
 FT TRANSMEM 898 917
 FT DOMAIN 918 923
 FT TRANSMEM 924 945
 FT DOMAIN 946 973
 FT TRANSMEM 974 992
 FT DOMAIN 993 1005

FT TRANSMEM 1006 1028 12 (BY SIMILARITY).
 FT DOMAIN 1029 1034 CYTOPLASMIC (BY SIMILARITY).
 SQ SEQUENCE 1034 AA; 111454 MW; B676807A5BD17B1 CRC64;
 Alignment Scores:
 Pred. No.: 0.00689 Length: 1034
 Score: 144.00 Matches: 135
 Percent Similarity: 31.73% Conserved: 89
 Best Local Similarity: 19.12% Mismatches: 226
 Query Match: 4.35% Indels: 256
 DB: Gaps: 30
 US-09-868-987-1 (1-1864) x ACRF_ECOLI (1-1034)
 QY 308 TTGGCAATGCTTATGTTTGGATGAGCGTATATTATGATTGGAGCGCATCGCTTCG 367
 DB 353 LeuValMetCysLeuPheLeuGlnAsnMetCysAlaThrLeuIleProThrIleAlaVal 372
 QY 368 GAGCTGTCTTGTGATCTTTGCTTATCTGGCGAGCTCTACAGTATTGGATGGCCA 427
 DB 373 ProValValLeuLeuGlyThrPheAlaIleLeuAlaIlePheGlyTyrSerIleAsnThr 392
 QY 428 CTCACCTGTTCAGACTCGCTGGGATTTCTTCTATGGGATGGCCGTAGATGCAAT 487
 DB 393 LeuThrMetPheGlyMet-----ValLeuAlaIleGlyLeuLeuValAspAspAla 409
 QY 488 GTTCTGTATTCCAGAAATCCAGAGAAATTTTATTGTCCTAAAGCTTT-----AAA 541
 DB 410 IleValValValGlnAsnVal---GluArgValMetMetGlnAspLysLeuProPheLys 428
 QY 542 AAATCTGTAGAAAGATATACCAAGCTTTGGACGATTTTGGATTTGATCTTAAGTACT 601
 DB 429 GlnAlaThrGlnLysSerMetCysGlnIleGlnAlaLeuVal-----GlyIleAla 446
 QY 602 ACAATATGGCTTCAGACTCTT-----TTCTCTCATGATACAGGCGCTATT 649
 DB 447 MetValLeuSerAlaValPheIleProMetAlaPhePheGlySerThrGlyAlaIle 466
 QY 650 ---AAAGCTTGTGCTTGAACATTTATAGAAATTTCTTCAATGTTTACGCTCTT 706
 DB 467 TyrArgGlnPheSerIleThrIleValSerAlaMetAlaLeuSerValLeuValAlaLeu 486
 QY 707 TTCATGACT----- 715
 DB 487 IleLeuThrProAlaLeuCysAlaThrLeuLeuLysProValSerAlaGlnHisGln 506
 QY 716 ---AAATTTTCTTCATGCTGTGGATGATAGAAC---CAACATACACAGTTGCATATG 769
 DB 507 AsnLysGlyGlyPhePheGlyTyrPheAsnThrThrPheAspHisSerValAsnHisTyr 526
 QY 770 ATGAAT-----AAGTTCTGGGGATTAAGCATATTTCTTG----- 805
 DB 527 ThrAsnSerValGlyLysIleLeuGlySerThrGlyArgTyrLeuLeuIleTyrAlaLeu 546
 QY 806 ---AGAGATGCCAAAAAATTTGGCGCTGTTCTGGAAGTCTTTCTT----- 850
 DB 547 IleValIleGlyMetValValLeuPheLeuArgLeuProSerSerPheLeuProGlnGln 566
 QY 851 ---TTAGTGGCTGCTCTCGCGTTGGAGCTTGGAAT----- 886
 DB 567 AspGlnIleValPheLeuThrMetIleGlnLeuProIleGlyAlaThrGlnArgThr 586
 QY 886 ----- 886
 DB 587 GlnLysValLeuAspGlnValThrAspTyrTyrLeuLysAsnGlnLysAlaAsnValGln 606
 QY 887 -----TCCGTTTGGGAATGATTTTAAAGA----- 913
 DB 607 SerValPheThrValAsnGlyPheSerPheSerGlyGlnIleAsnAlaGlyMetAla 626
 QY 913 ----- 913
 DB 627 PheValSerLeuLysProTyrGlnGlnArgAsnGlyAspGlnAsnSerAlaGlnAlaVal 646

QY 914 -----GGGTATGCGCTTTACCTTTAAAT 934
Db 647 ILeHisArgAlaLysMetGluLeuGlyLysIleArgAspGlyPheValIleProPheAsn 666
QY 934 ----- 934
Db 667 MetProAlaIleValGluLeuGlyThrAlaThrGlyPheAspPheGluLeuIleAspGln 686
QY 934 ----- 934
Db 687 AlaGlyLeuGlyHisAspAlaLeuThrGlnAlaArgAsnGlnLeuLeuGlyMetAlaAla 706
QY 935 -----CRAAAGAGCATGCATCAGCGATGTTGCTCAATG 970
Db 707 GlnHisProAlaSerLeuValSerValArgProAsnGlyLeuGluAspThrAlaGlnPhe 726
QY 971 CGTGGCAAGTT---GTGCATAAACTACAGAGAGCTGCTCTTCTCTAGAGACTTCGT 1027
Db 727 LysLeuGluValAspGlnGluLysAlaGlnAlaLeuGlyValSerLeuSerAspIleAsn 746
QY 1028 ATTCAACATTGGATCTCTCA----- 1048
Db 747 ---GlnThrIleSerThrAlaLeuGlyThrTyrrValAsnAspPheIleAspArgGly 765
QY 1049 -----GAAAGATCAAAATCTATTTAGTGATAAA 1078
Db 766 ArgValLysLysLeuTyrrValGlnAlaAspAlaLysPheArgMetLeuProGluAspVal 785
QY 1079 GCTTTAAGCTATCTAGCAGATACGACCTCTCTCTCA-----AAATTACGATC 1129
Db 786 AspLysLeuTyrrValArgSerAlaAsnGlyGluMetValProPheSerAlaPheThrThr 805
QY 1130 ATGAGCTGG-----CGTATTGTGGGATT---GTTGTGAGA 1162
Db 806 SerHisTrpValTyrrGlySerProArgLeuGluArgTyrrAsnGlyLeuProSerMetGlu 825
QY 1163 AACAGGCTAGATTCTACGAAACTCTAAACGAACGCA---AAATTGTGTCAAAG 1219
Db 826 ILeGlnGlyGluAlaAlaProGlyThrSerSerGlyAspAlaMetAlaLeuMetGluAsn 845
QY 1220 GTAAGCAGCAACTATCG-----AGAAATCGTTCATCGCG 1258
Db 846 LeuAlaSerLysLeuProAlaGlyIleGlyTyrrAspTrpThrGlyMetSerTyrrGlnGlu 865
QY 1259 ACCATCGGG-----CTTTAGAGCTTTGGCAATCATCTGCTCTATGTG 1303
Db 866 ArgLeuSerGlyAsnGlnAlaProAlaLeuValAlaIleSerPheValValValPheLeu 885
QY 1304 AGTTTGGCTTTGAATGGAATATGCTTTTCAGTCCGATCGCTTTAATTCATGACCTT 1363
Db 886 CysLeuAlaAlaLeuTyrrGluSerTrpSerIleProValSerValMetLeuValValPro 905
QY 1364 TTGCTACCTGTGCAGCTCTGTTTATACCATTTCTTTTGGAGAAATTCAAATAGAT 1423
Db 906 LeuGlyIleValGlyValLeuLeuAlaThrLeuPheAsnGlnLys-----AsnAsp 923
QY 1424 TTGCAAGCCATTGTGCTTTAATGACTGTATGGGTATTCTATTAACAATCTTTGATC 1483
Db 924 ValTyrrPheMetValGlyLeuLeuThrIleGlyLeuSerAlaLysAsnAlaIleLeu 943
QY 1484 ATTTTTCATCGTATTCGGAAGATGCGCAAGCAACCTGTTTACCCCTATGATGTTT 1543
Db 944 ILeValGlu-----Phe 947
QY 1544 GTTAATGATCCCTTCAAAAG-----ACGTTACGCCGACGGTAATGACACA 1591
Db 948 AlaLysAspLeuMetGluLysGluGlyValValGluAlaThrLeuMetAlaVal 967
QY 1592 GCTCAAACTCTACGATTTCTGTTAATG-----CTTTGTTTATA----- 1630
Db 968 ArgMetArgLeuArgProIleLeuMetThrSerLeuAlaPheIleLeuGlyValLeuPro 987

QY 1631 -----GGGGCTCTCTGCTCTTTAAATTTTGCATTTATATGACCATA 1672
Db 988 LeuAlaIleSerAsnGlyAlaGlySerGlyAlaGlnAsn-----AlaValGlyIle 1004
QY 1673 GGGATTCTCTAGAACTTTATCGTCTCTTTATATTGCA-----CCACCTCTGTTG 1723
Db 1005 GlyValMetGlyMetValSerAlaThrLeuLeuAlaIlePhePheValProValPhe 1024
QY 1724 TTGTTTATGGTCGCTAA 1741
Db 1025 PheValValIleArgArg 1030
RESULT 34
Y895 HAEIN STANDARD; PRT; 1032 AA.
AC Q57124; O05034;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein HI0895.
GN HI0895.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
SEQUENCE FROM N.A.
RP STRAIN=Rd / KW20 / ATCC 51907;
RC MEDLINE=95350630; PubMed=7542800;
RX Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA McLavigne A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fierle B., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.B., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
Rd.";
RL Science 269:496-512 (1995).
CC -!- FUNCTION: COULD BE A DRUG EFFLUX PUMP (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
(By similarity).
CC -!- SIMILARITY: BELONGS TO THE ACRB/ACRD/ACRF FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U32771; AAC22555.1; -;
DR TIGR; HI0895; -;
DR InterPro; IPR001036; Acrflvin_res.
DR Pfam; PF00873; ACR tran; 1.
DR PRINTS; PR00702; ACRFLVINRNP.
KW Hypothetical protein; Transmembrane; Inner membrane; Transport;
KW Complete proteome.
FT DOMAIN 1 17 CYTOPLASMIC (BY SIMILARITY).
FT TRANSMEM 18 36 1 (BY SIMILARITY).
FT DOMAIN 37 337 PERIPLASMIC (BY SIMILARITY).
FT TRANSMEM 338 357 2 (BY SIMILARITY).
FT DOMAIN 358 363 CYTOPLASMIC (BY SIMILARITY).
FT TRANSMEM 364 383 3 (BY SIMILARITY).
FT DOMAIN 384 389 PERIPLASMIC (BY SIMILARITY).
FT TRANSMEM 390 411 4 (BY SIMILARITY).
FT DOMAIN 412 438 CYTOPLASMIC (BY SIMILARITY).
FT TRANSMEM 439 457 5 (BY SIMILARITY).
FT DOMAIN 458 470 PERIPLASMIC (BY SIMILARITY).
FT TRANSMEM 471 493 6 (BY SIMILARITY).

Db 830 AsnLeuProGlnGlyTyThrPheAspPheLysSerGluAlaArgGlnLeuValGlnGlu 849
 Qy 1262 -----ATCGGGCTTTTAGGAGCTTTGGCAATCACTTTGCTCTATGCTAGTGGTGGC 1312
 Db 850 GlyAsnAlaLeuAlaValThrPheAlaLeuAlaValIleIlePheLeuValLeuAla 869
 Qy 1313 TTGAATGGCAATATGCTTTTCAGTGGCGTATGCTGCTTTAATTCATGACCTTTTGGCTACC 1372
 Db 870 IleGlnPheGluSerIleArgAspProMetValIleMetIleSerValProLeuAlaVal 889
 Qy 1373 TGTGCGAGCTTTG-----TTATAGCACATTTCTTTTGAAGAAATCAATA 1420
 Db 890 SerGlyAlaLeuValSerLeuAsnIleLeuSerPhePheSerIleAlaGlyThrThrLeu 909
 Qy 1421 GATTGCAAGCATTGGTCTTTAATGACTGTATTTGGGTATTCATTAAACAATCTTTG 1480
 Db 910 AsnIleTySerGlnValGlyLeuIleThrLeuValGlyLeuIleThrLysHisGlyIle 929
 Qy 1481 ATCATTTTGTATGCTATTCGTAAGATCGCCAGCAACCTGTTTACCCCTATGATGTT 1540
 Db 930 LeuMetCysGluValAlaLysGluGlnLeuAsnHisGlyLysThrArgIleGluAla 949
 Qy 1541 TTAGTTAATGATGCTCTCAAAAGACGTTACGCCACGGTAATGACACACAGCTACAACT 1600
 Db 950 IleThrHis---AlaAlaLysValArgLeuArgProIleLeuMetThrThrAlaAlaMet 968
 Qy 1601 CTATCAGTTTGTAAATGCTTTGTTTATAGCGGCTCC-----TCGTCTTTAAAT 1651
 Db 969 ValAlaGlyLeuIleProLeuLeuTyThrAlaThrGlyAlaGlyAlaValSerArgPheSer 988
 Qy 1652 TTTCATTTATATGACCATAGGATTCCTTCTAGCACTTATGCTCTCTTTATATGCA 1711
 Db 989 IleGlyIleValIleValAlaGlyLeuSerIleGlyThrIlePheThrLeuPheValLeu 1008
 Qy 1712 CCA-----CCTCTGTTGTTGTTGTTTATGTCGTT 1738
 Db 1009 ProValValTySerTyThrValAlaThrGluHisLysProLeuProValPhe----- 1025
 Qy 1739 AAAGAAATCGCTCA 1753
 Db 1026 AspGluAsnLysThr 1030

RESULT 35
 MMLB STRCO
 ID MMLB STRCO STANDARD; PRT; 847 AA.
 AC OS4101;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DE 15-JUN-2002 (Rel. 41, Last annotation update)
 GN Putative membrane protein SC05905.
 OS SC05905 OR SC10A5.10C.
 GN Streptomyces coelicolor.
 OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
 OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
 OX NCBI_TaxID=1902;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2) / M145;
 RX MEDLINE=21996410; PubMed=12000953;
 RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
 RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
 RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
 RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
 RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neil S.,
 RA Rabinowitz S.E., Rajandream M.A., Rutherford K., Rutter S.,
 RA Seger K., Saunders D., Sharp S., Squares R., Squares R., Taylor K.,
 RA Warren T., Wietzorrek A., Woodward J., Barrrell B.G., Parkhill J.,
 RA Hopwood D.A.;
 RT "Complete genome sequence of the model actinomycete Streptomyces
 coelicolor A3(2).";
 RL Nature 417:141-147(2002).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).

CC -1- SIMILARITY: BELONGS TO THE MMLB FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AL021529; CAA16442.1; -;
 DR InterPro; IPR000731; HMGR/patch_5TM.
 DR InterPro; IPR004889; MMLB.
 DR Pfam; PF03176; MMLB; 2.
 DR PROSITE; PS0156; SSD; 2.
 KW Hypothetical protein; Transmembrane; Complete proteome.
 FT TRANSMEM 18 38 POTENTIAL.
 FT TRANSMEM 187 207 POTENTIAL.
 FT TRANSMEM 215 235 POTENTIAL.
 FT TRANSMEM 248 268 POTENTIAL.
 FT TRANSMEM 302 322 POTENTIAL.
 FT TRANSMEM 326 346 POTENTIAL.
 FT TRANSMEM 381 401 POTENTIAL.
 FT TRANSMEM 539 559 POTENTIAL.
 FT TRANSMEM 562 582 POTENTIAL.
 FT TRANSMEM 600 620 POTENTIAL.
 FT TRANSMEM 643 663 POTENTIAL.
 FT TRANSMEM 672 692 POTENTIAL.
 FT TRANSMEM 744 764 POTENTIAL.
 FT TRANSMEM 778 798 POTENTIAL.
 FT TRANSMEM 808 828 POTENTIAL.
 SQ SEQUENCE 847 AA; 88616 MW; E2A443F5C74AE1DE CRC64;
 Alignment Scores:
 Pred. No.: 0.00358 Length: 847
 Score: 135.00 Matches: 123
 Percent Similarity: 33.33% Conservative: 80
 Best Local Similarity: 20.20% Mismatches: 244
 Query Match: 4.08% Indels: 162
 DB: 1 Gaps: 23
 US-09-868-987-1 (1-1864) x MMLB_STRCO (1-847)
 Qy 173 GAAGTAGCAAACTCGCTCTCAGATTTAAATCTGGACGATGCTTTTGT-CCC 226
 Db 155 GluValArgAlaLeuAlaSerArgAlaGlyProAlaValAlaValThrGlyPro 174
 Qy 227 GAGGTTCTCAGTAGAGAGAGATCTCTTCTGATCTTGGGAAAAACAATGTACACAGGC 286
 Db 175 AlaGlyIleAlaThrAspThrValLysValPheSerGlyGlyAspLys-190
 Qy 287 ATTATCTCAGCATGCTGCTGCTTGGCAATGCTTATTTGATGAGCGTATATTATAGA 346
 Db 191 ValLeuLeuAlaThrValValLeuValLeuIleLeuLeuAlaIleTyArg-209
 Qy 347 TTTGAGGCGTCATCGCTTCGGGAGCTGTTCTTCTGAATCTTTTGTATCTTATCTGGCAGCT 406
 Db 210 -----SerProLeuMetAlaLeuValProLeuLeuAlaValGlyValAla 224
 Qy 407 CTACAGTATTTGGATGCGCCACTCCTTGTGAGGACTCGCTGGGATT-454
 Db 225 MetArgValAlaGluThrLeuGlyAlaIleLeuAlaAspAlaGlyValIleThrValSer 244
 Qy 455 -----GTTCTTGTATGCGGATGCGCGTAGATGCAATGTT 490
 Db 245 SerGlnThrAlaSerIleMetThrValLeuLeuPheGlyValGlyThrAspTyThrAlaLeu 264
 Qy 491 CTGTATTCGAAAGAACCCGAGAGAAATTTTATTTGTCTCAAGTCTTAAATAATCTGTA 550
 Db 265 IleIleThrAlaArgTyArgGluThrLeuLeuAspGluProAspArgAlaAlaMet 284
 Qy 551 GAAAGAGATATACCAAGGCTTTTGGAGCCATTTTGAATCTTAAGTCTTACACTGACAGTATTG 610

```

Db 285 GlnAlaIaIaValArgArgThrAlaGluSerValIleuAlaSerAlaSerThrIleValIleu 304
Qy 611 GCCTGACACTTCTTTCTTCCTAGATACAGGGCCTATTAAAGGTTGCTTGACATTG 670
Db 305 AlamecPheAlaLeuIleValAlaIaValSerProAlaLeuHisGlyPheGlyProTyrIleu 324
Qy 671 ATTTAGAAATTTCTCTCAATGTTTACGGCTCTTTCACGACATAATTTTCTTCATG 730
Db 325 AlaIeuGlyValAlaIaValMetAlaLeuValAlaPheThrPheIleProAlaLeuValIleu 344
Qy 731 CTGTGATGAATGAAGACCCAACTACAGTTGATATGATGAAATGAATGCTGGGGGATA 790
Db 345 Leu----- 345
Qy 791 AAGCATGATTTCTTGAGAGATGCAAAAACTTGGGCTGTTTCTGGAAGTGTCTTCTT 850
Db 346 -----LeuGlyArgSerValPheTyr 352
Qy 851 TTAGTTCGCTT-----GCTCTGGGTTTGAGGCTTGGAATTCGCTT 892
Db 353 ProGlyGlyValAspIlyAlaIaGluArgSerArgIlyAlaGlyIleTyrHisArgIle 372
Qy 893 TTGGGATGATTTTAAAGA----- 913
Db 373 AlaAlaLeuValAlaIaArgAlaProValIlyValAlaSerAlaValIleAlaLeuVal 392
Qy 914 -----GGGTATGCTTTTACCTTTAAT----- 934
Db 393 ValIeuSerAlaGlyLeuIleuGlyIlyrGlnIleuSerPheAsnThrLeuSerGlyPheArg 412
Qy 935 -----CCAAAGACATGCGCATCGCATGTCGTCGAATGCCGGGCAA----- 979
Db 413 AlaAlaThrGluSerGluHisGlyIleHisLeuIleArgGluGluPheGlyProGlyIleu 432
Qy 980 -----GTTGTCATTAACCTACAGAGAAGCTGCTTCTTCTTACAGAC 1021
Db 433 IleAlaProSerThrValValValHisSerGlnAspAsnLeuArgSerSerProAlaPro 452
Qy 1022 TTCCGATTTCAACA----- 1036
Db 453 AlaAspIleAlaThrAlaLeuThrAspAlaAspHisValSerArgValAlaAspProArg 472
Qy 1037 TTTGGATCTTCAAGAAATCAATAATCTAT-----TTTACGT 1072
Db 473 MetGlyAspGlyIlyrThrValPheTyrAspValIleLeuAspLeuAspProTyrSer 492
Qy 1073 GATAAGCTTTAAGCTATACAGATACAGAGCTCTCTCCCTAAATTAACGATCATG 1132
Db 493 SerIlyAlaLeuAspAlaIle-----GlyProIleuIlyGlnAlaThrGlnSer 508
Qy 1133 AGCTGCGCTTATTTGGGATTGTTGTCAAAAACAGGCTAGATTTCTACGGAACCTCT 1192
Db 509 AlaIaGlnAlaIaGly-----ValGlnAspAlaThrValIleuIleGlyGlyIleuThr 526
Qy 1193 AAACGAAACGCAAAATTTGGTCAAGGTAAAGCAGCAACTATGCAAGAAATGCGTTAT 1252
Db 527 AlaGlnAsnAlaIa-----AspIleArgSerAlaLeuAspArg----- 538
Qy 1253 CAGGAGCAATCGGCTTTTGGAGCTTTGGCAATCATCTGCTGTATGATGAGTTGGCGC 1312
Db 539 AspThrThrIleuIleValIleuValIleuAlaIleValIleuValIleuValIleu 558
Qy 1313 TTTGAATGCAATATGCTTTGACGCGCTAGCGCTTTAATTCATGACCTTTTGGCTACC 1372
Db 559 LeuArgSerIleuLeuAlaIleProIeuTyrIleuValAlaIaThrIleu-----LeuIleSerPhe 576
Qy 1373 TGTGACGCTTCTTTATGACACATTTCTTTTGAAGAAATTCAAATGATTTGCAAGCC 1432
Db 577 LeuAlaThrIleuGlyAlaIleThrThrPhePheThrValThrValIleuGlyAsp---AspGly 595
Qy 1433 ATTTGCT-----GCTTAATGACGTATGAGGGGATTCATTA 1468
Db 596 IleGlyAsnArgValThrAlaTyrIlePheValIleuValAlaIleuGlyValAlaAspTyr 615

```

```

Qy 1469 AACAAATCTTGATCATTTTGTGATCGTATTCGTGAAGATCCCAAGCAACCTGTTTACC 1528
Db 616 Asn-----IlePheIleMetSerArgPheIlyGlnIleuArgThrGln----- 630
Qy 1529 CCTATGATGCTTTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1579
Db 631 ProProAlaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 650
Qy 1580 GTAATGCAACAGCTACCACTATCATGTTTGTGATATGCTTTGTTTATGAGCGGCTCC 1639
Db 651 GlyIleuIleuAlaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 667
Qy 1640 TCTGCTTTTATTTTGTGATTTATTTATGACATATGAGGATTTCTTACGAACTTTATGCTCT 1699
Db 668 GluIleuPheGlnPheGlyPheAlaMetAlaCysGlyIleuLeuAspThr----- 684
Qy 1700 CTTTATATGACACACCTGTTGTTG 1726
Db 685 ---PheIleuIleArgProIleuVal 692

RESULT 36
YC53 METUA STANDARD; PRT; 282 AA.
AC 058650;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein MJ1253.
GN MJ1253.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococcaceae; Methanocaldococcus.
ON NCBI_TaxID=2190;
RX STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
MEDLINE=9637999; PubMed=8688087;
RA Bait C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrman J.L., Nguyen D.,
RA Utermack T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kane B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii."
RL Science 273:1058-1073(1996).
CC -|- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -|- SIMILARITY: BELONGS TO THE SECY/SECF FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U67565; AAB99256.1; -.
CC TIGR: MJ1253; -.
CC InterPro: IPR003335; SecD_SecF.
CC Pfam: PF02355; SecD_SecF.1.
CC Hypothetical protein; Protein transport; Translocation; Transmembrane;
CC Complete proteome.
CC
CC TRANSMEM 9 29 POTENTIAL.
CC TRANSMEM 120 140 POTENTIAL.
CC TRANSMEM 149 169 POTENTIAL.
CC TRANSMEM 174 194 POTENTIAL.
CC TRANSMEM 214 234 POTENTIAL.
CC TRANSMEM 236 256 POTENTIAL.
CC SEQUENCE 282 AA; 31029 MW; D2AC5859AEFC1079 CRC64;

```

[illegible]

Oy	356	GTCAATCGTTGGAGGAGCTTTCTTCTTGAATCTTTCTTATCTGGGACGCTACAGAT	415
Db	369	ITelleProGlyValAlaIeValProLeuSerLeuIIleGlyThrPheAlaValMetAlphe	388
Oy	416	TTGAGTGGCCACTCACCTTGTGCAGACCTGCCTGGAGATTGTTCTTGCTATGGGAGGCC	475
Db	389	LeuAspPheSerIleAsnLeuMetThrLeuMetAlaLeuThrIleAlaThrGlyPheVal	408
Oy	476	GTAGATGCMAATGTTCTTGATTCGAAAGAAATC-----CGAGAGCAATT	520
Db	409	ValAspAspAlaIleValIleGluAsnIleSerArgTyrIleGluIleSelyGluIlys	428
Oy	521	TTATGTCGCAAGCTTTAAAAAACTGTGTAAGAAAAGATATACCAAGGCTTTGGAGCC	580
Db	429	ProLeuAlaAlaIleLeuIysGlyAlaGlyIleIleGlyPheThrIleIle-----	445
Oy	581	ATTTTGTATTTAACTTAACTTACAGATATGGCTTCAGACCTCTTCTTCTCTAGATACA	640
Db	446	-----SerLeuThrPheSerLeuIleAlaValLeuIleProLeuMetPheMet	461
Oy	641	GGGCGCTATTAAGG-----TTTGCCTTGACATTTAGAAATTTC	685
Db	462	GlyAspIleValGlyArgLeuPheArgGluPheAlaIleThrLeuAlaValAlaIleLeu	481
Oy	686	TCTTCATAGTTTACGGCTTTTTCATACACAAATTTTCTCATGCTGGTGATGAATAG	745
Db	482	IleSerAlaValIleSerLeuThrLeuThrPro-----MetCysAlaArgMet	498
Oy	746	ACCAACATACACAGTTCGCATATGATGAATGAATGTCGGGAGTAAGCATGATTTCTTG	805
Db	499	LeuSerGlnGluSerLeuAlaGlyGlnAsnArgPheSerArgAlaSerGluIysMetPhe	518
Oy	806	-----AGAGATGCMAAAACTT-----TGCGCT	829
Db	519	AspArgIleIleAlaIleTyrGlyArgGlyLeuAlaIleValLeuAsnIleProTyr-Le	538
Oy	830	GTTTCTGGAATGTTTTCTTTTAGTGTGGCTGCCTCGGCTTTGGAGCCGGAATTC	889
Db	538	uThrLeuSerVal-----AlaLeuSerThrLeuLeuLeuSerValLe	552
Oy	890	GTTTGGGAATGATTTTAAAGAGGGTATGCTTTACTT-----TTATTCACAA	940
Db	552	uLeuThrValPheIleProIysGlyPhePheProValGlnAspAsnGlyIleIleGlnGlnI	572
Oy	941	GAGCATGCATCAGCGATGTTGCTCAAAATGCTGGCAAAATGTGCTATTAACATCAGAA	1000
Db	572	yThrIleuGlnAlaPro-----	577
Oy	1001	GCTGCTCTTCTTCTAGAGACTCCGATTCACAACTTTGATCTTCACAAAAGATCAA	1066
Db	578	-----GlnSerSerPheAlaAsnMetAlaGlnArgGlnValGlnVal	592
Oy	1061	ATCTATTTAGTGAATTAAGCTTTAGCTTATCTATACACAGATACAGAGCTCTCTCTAA	1120
Db	592	IAlaAspValIleLeuGlnAspProAlaValGlnSer-----	604
Oy	1121	TTAACGATCATGAGCTGCGCTTATTGAGATTGTTGTACAGAAACGGCTAGATTCTC	1180
Db	605	-----LeuThrSerPheValGlyValAspGlyTh	614
Oy	1181	TACGAAACTTAAACG-----AACGAAATTTTGGTCAAAAGTAAAG	1224
Db	614	AsnProSerLeuAsnSerAlaArgLeuGlnIleAsnLeuIysProLeuAspGlnArgAs	634
Oy	1225	CAGCAAACTATCGAAGAAATG---GGTATFACGGGACGATCGGGCTTTTGGAGCTTT	1281
Db	634	AspAspArgValGlnIysValIleAlaArgLeuGlnThrAlaValAspIysValProGlyVa	654
Oy	1282	GGCACTATCTTTG-----CTCTATGATGATTTGCGCTT	1314
Db	654	IAspLeuPheLeuGlnProThrGlnAspLeuThrIleAspThrGlnValSer---ArgTh	673
Oy	1315	TGAATGCGAATATGCTTTCAGTCCGCTATGCGCTTTAATTCATGACCTTTTGGTAACTCG	1374

DR InterPro; IPR000731; HMGCR/patch_5TM.
 DR InterPro; IPR004869; MMPL.
 DR Pfam; PF03176; MMPL; 1.
 DR PROSITE; PS50156; SSD; 1.
 KW Hypothetical protein; Transmembrane; Complete proteome.

FT TRANSMEM 14 34 POTENTIAL.
 FT TRANSMEM 186 206 POTENTIAL.
 FT TRANSMEM 210 230 POTENTIAL.
 FT TRANSMEM 236 256 POTENTIAL.
 FT TRANSMEM 287 307 POTENTIAL.
 FT TRANSMEM 315 335 POTENTIAL.
 FT TRANSMEM 397 417 POTENTIAL.
 FT TRANSMEM 563 583 POTENTIAL.
 FT TRANSMEM 587 607 POTENTIAL.
 FT TRANSMEM 617 637 POTENTIAL.
 FT TRANSMEM 673 693 POTENTIAL.
 FT TRANSMEM 699 719 POTENTIAL.
 FT CONFLICT 756 756 K -> N (IN REF. 2).
 SQ SEQUENCE 944 AA; 100904 MW; 289DF83ACDBE490D CRC64;

Alignment Scores:

Pred. No.: 0.0133 Length: 944
 Score: 128.00 Matches: 126
 Percent Similarity: 36.01% Conservative: 121
 Best Local Similarity: 18.37% Mismatches: 238
 Query Match: 3.86% Indels: 201
 DB: 1 Gaps: 31

US-09-868-987-1 (1-1864) x MML3_MYCTU (1-944)

Qy 68 GGATGGCGTATGGCTGTAGTATTGACGGTATATATGGTCAGCAGC----- 112
 Db 107 GlyTrrAla-----GlyTrrLeuAArgAlaSerGlnAlaThrGlyMet 120
 Qy 113 -----CCTATTTTAAACGTCCTCATTAAGAA----- 136
 Db 121 AlaThrAlaAspLysLysTyrrThrPheValSerIleProLeuLysGlyAspAspAsp 140
 Qy 137 -----AATCATGCCAGTGTCTCAGGGAATTTTACCACCGTGAAGTGAACAA 184
 Db 141 ThrIleLeuAsnAsnTyrrLysAlaIleAla-----ProAspLeuGlnArg 155
 Qy 185 CTC---GCCTCAGATTTAAATCTGGACGATGCTTTTGTTCGCGAGTCTTCAGTAA 241
 Db 156 LeuAspGlyGlyThrValLysLeuAlaGlyLeuGlnProValAlaGluAlaLeuThr-- 174
 Qy 242 GAGACGATCTCTTGATCTTGGGAAAAACAATGTACACAGGCATTTCTCAGCATGC 301
 Db 175 GlyThrIleAlaThrAspGlnArgMetGluVal----- 186
 Qy 302 TGTGGCTTGGCAATGCTTATTTGTATGAGCGTATATTAGATTTGGAGCGGTATC 361
 Db 187 -----LeuAlaLeuProLeuValAlaValValLeuPheValPheGlyGlyValIle 204
 Qy 362 CTTTCGGAGCTGTT-----CTTCGT 382
 Db 205 AlaAlaGlyLeuProValMetValGlyGlyLeuCysIleAlaGlyAlaLeuGlyIleMet 224
 Qy 383 AATCTTTTGTCTATCTGGCAGCTCTACAGTATTTGGATGGCCACTCACCTTGTCA 442
 Db 225 ArgPheLeuAlaIlePheGlyProValHisTyrrPheAlaGlnProValVal----- 241
 Qy 443 CTCGGTGGATTTGTTGCTATGGGATGGCGGTAGATGCAAAATGTTCTTCTATTG 502
 Db 242 -----SerLeuIleGlyLeuGlyIleAlaIleAspTyrrGlyLeuPheIleValSer 258
 Qy 503 AGAATCCGAGAGGAATTTTATTTGCTCAAGTCTTAAATAATCTGTAGAAAAA----- 556
 Db 259 ArgPheArgGluGluIleAlaGluGlyTyrrAspThrGluThrAlaValArgArgThrVal 278
 Qy 557 -----GGATATACCAAGCTTTTGAGCCATTTTGTATCTTAATCTGACTACATA 607
 Db 279 IleThrAlaGlyArgThrValThrPheSerAlaValLeu-----Ile 292

Qy 608 TTGCCTCAGCA-----CTTCTTTTCTCTAGATACAGGCGCTATTAAAGGTTTCT 661
 Db 293 ValAlaSerAlaIleGlyLeuLeuPhe-----ProGlnGlyPheLeu 307
 Qy 662 TTGACATTGATTTAGGAATTTCTCTCAATGTTTACGGCTCTTTTCATCACTAATTT 721
 Db 308 LysSerLeuThrTyrrAlaThrIleAlaSerValMetLeuSerAlaIleLeuSer----- 325
 Qy 722 TTCTTCATGCTGTGGATGAATAAGACCAACATACACAGTTGTCATATGATGAATAAGTTC 781
 Db 326 -----IleThrValLeuProAlaCys 332
 Qy 782 GTGGGATA-----AAGCATGATTTCTTGAGAGATGCAAAAACTT----- 823
 Db 333 LeuGlyIleLeuGlyLysHisValAspAlaLeuGlyValArgThrLeuPheArgValPro 352
 Qy 824 -----TGGCGTGTCTTCTGGAAGTGTCTTTCTTTAGTGGTGGTGGCT----- 865
 Db 353 PheLeuAlaAsnTrpLysIleSerAlaAla-----TyrLeuAsnTrpLeuAlaAspArg 370
 Qy 866 -----CTCGGTTTGGAGCTGCAATTCCTGTTTGGGA 898
 Db 371 LeuGlnArgThrLysThrArgGluValGluAlaGlyPheTrpGlyLysLeuValAsn 390
 Qy 899 ATGGATTTTAAA----- 910
 Db 391 ArgValMetLysArgProValLeuPheAlaAlaProIleValIleIleMetIleLeuLeu 410
 Qy 911 -----GGAGGTATGCTTTTACCTTTTAAATCCAAAGAG 943
 Db 411 IleIleProValGlyLysLeuSerLeuGlyIleSerGluLysTyrrLeuProThr 430
 Qy 944 CATGCATCAGCATGTGCTCAATCGTCGCAAGTGTGCATAAACTACAGC----- 998
 Db 431 AspSerValArgGlnAlaGlnGluGlu-PheAspLysLeuPheProGlyTyrrArgThrAs 450
 Qy 999 -AAGCTGTCTTCTTCTAGAGCTTCCGTATTCAAAACATTG--GATCTTCAG----- 1049
 Db 450 nProLeuThrLeuValIleGlnThrSerAsnHisGlnProValThrAspAlaGlnIleAl 470
 Qy 1050 -----AAAAGATCAAAATCTATTATGATGATAAAGCTTT----- 1083
 Db 470 aAspIleArgSerLysAlaMetAlaIle-GlyGlyPheIleGluProAspAsnAspProa 490
 Qy 1084 -----AAGCTATACTAAGCAGATACAGCGCTCTCTCTAAATTAACG 1126
 Db 490 laAsnMetTrpGlnGluArgAlaTyrrAlaValAlaSerLysAspProSerValArgV 510
 Qy 1127 ATCATGAGCTGGCGTATTGTGGGATTGTTGCAGAACAGCGCTAGATTTCTCTACGGA 1186
 Db 510 alLeuGlnAsnGlyLeuIleAsnProAlaAspAlaSerLysLysLeuThrGluLeuArgA 530
 Qy 1187 AACTCTAAACGAAACGCAAAATTTTGGTCT-----AAAGGTAAAGCAGCAAACTATCGAAGAAA 1243
 Db 530 laIleThrProLysGlyIleThrValLeuValGlyGlyThrProAlaLeuGluLeuA 550
 Qy 1244 ATGGTTATCAGGCGACCATCGGCTTTTAGAGCTTTGGCAATCATCTTCTCTATGTG 1303
 Db 550 spSerIleHisGly-LeuPheAlaLysMetProLeuMetValValIleLeuLeuThrThr 569
 Qy 1304 AGTTTGGCTTTGAATGGCAATATGCTTTTACGTCGCGTATCGCTTTTAAATTCATGACCTT 1363
 Db 570 ThrIleValLeuMetPhe---LeuAlaPheGlySerValValLeuProIleLysAlaThr 588
 Qy 1364 TTGGTACCTGTGCAGTCTTG-----TTTATAGCA 1393
 Db 589 LeuMetSerAlaLeuThrLeuGlySerThrMetGlyIleLeuThrTrpIlePheValAsp 608
 Qy 1394 CATTTCTTTTGAAGAAATTCAA-----ATAGATTTTCAAGCCCATTTGGT 1438
 Db 609 GlyHisPheSerLysTrpLeuAsnPheThrProThrProLeuThrAlaProValIleGly 628

QY 1439 GCTTTATGACT---GTAITGGGGTATTCATTAAACAATCTTGATCATTTTGATCGT 1495
 Db 629 Leuilellelealeuvalpneglyleuserthrasyprylgluvalphelevalserarg 648
 QY 1496 ATTGCTAAGATCGCCAA3CGAACCTGTTTACCCCTTATGCATGTTTATGATGATGCC 1555
 Db 649 Metvalgluvalaargluarglymetserthringleuvalaileargilegly----- 666
 QY 1556 CTTCAAAAAGCGTTCAGCCGCAAGCTGTAATGACAACAGCTACACTATCATCTTTGTTA 1615
 Db 667 -----ThrAlaAlaThrGlyArgIleIleThrAlaAlaAlaLeu 679
 QY 1616 ATGCTTTTGTATAGCGCGCTCCTGCTGTTTAAAT-----TTT 1654
 Db 680 ILevalaIaValaIaIaGlyAlaPheValaPheSeraspLeuValMetMetLysTyrLeu 699
 QY 1655 GCAATTATATAGACCATAGGATCTCTTA---GGAACCTTATGCTCTTATATATGCA 1711
 Db 700 AlaPheGlyLeuMetAlaAlaAlaLeuLeuAspAlaThrValaValaArgMetPheLeuVal 719
 QY 1712 CCACCTCTGTTG 1723
 Db 720 ProserValMet 723
 RESULT 39
 MMJ3_MYCLE
 ID MMJ3_MYCLE STANDARD; PRT; 955 AA.
 AC 006081;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Putative membrane protein mmj3.
 GN MMJ3 OR ML2620 OR MLC1622.18C.
 OS Mycobacterium lepreae.
 OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 NC NCB1_TaxID=1769;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=TW;
 RX MEDLINE=2118732; PubMed=11234002;
 RA Cole S.T., Eigmeier K., Parkhill J., James K.D., Thomson N.R.,
 RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
 RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R.M., Devlin K., Duthoy S., Fellwell T., Fraser A., Hamlin N.,
 RA Holroyd S., Hornsby T., Jagsels K., Lacroix C., Maclean J., Moule S.,
 RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
 RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
 RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
 RA Barrett B.G.;
 RT Massive gene decay in the leprosy bacillus.";
 RL Nature 409:1007-1011(2001).
 CC - SUBCELLULAR LOCATION: Integral membrane protein (Potential).
 CC - SIMILARITY: BELONGS TO THE MMPL FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: Z95398; CAB08813.1; -;
 CC EMBL: AL583926; CAC32152.1; -;
 DR Leptoma; ML2620; -;
 DR InterPro: IPR004869; MMPL.
 DR Pfam: PF03176; MMPL; 1.
 KW Hypothetical protein; Transmembrane; Complete proteome.
 FT TRANSMEM 14 34 POTENTIAL.
 FT TRANSMEM 191 211 POTENTIAL.
 FT TRANSMEM 231 251 POTENTIAL.
 FT TRANSMEM 292 312 POTENTIAL.

FT TRANSMEM 320 340 POTENTIAL.
 FT TRANSMEM 402 422 POTENTIAL.
 FT TRANSMEM 568 588 POTENTIAL.
 FT TRANSMEM 592 612 POTENTIAL.
 FT TRANSMEM 622 642 POTENTIAL.
 FT TRANSMEM 679 699 POTENTIAL.
 FT TRANSMEM 704 724 POTENTIAL.
 SQ SEQUENCE 955 AA; 102792 MW; 8AP86562EB91D07B CRC64;
 Alignment Scores:
 Pred. No.: 0.0212 Length: 955
 Score: 125.50 Matches: 135
 Percent Similarity: 35.09% Conservative: 112
 Best Local Similarity: 19.18% Mismatches: 244
 Query Match: 3.79% Indels: 213
 Gaps: 29
 US-09-868-987-1 (1-1864) x MMJ3_MYCLE (1-955)
 QY 50 CAATATTCGCAAAACCGT-----GATGCGTATGCGCTGATGATTGAC 94
 Db 96 GlnPheGlnArgAsnAsnSerAsnLysValThrGlyTyrAla----- 109
 QY 95 GGTATATGATGTCAGACGCCCTATTTAAAGCTGCAATGAAAATCATGCGCAGTGTCTCA 154
 Db 110 GlyTyrLeuArgAlaSerAspThrThrAsnThrValaValaGlnGlyMetAlaThrProAsp 129
 QY 155 GGGAAATTTACCCACCGTGAAGTG----- 178
 Db 130 LysLysTyrThrPheValSerIleProLeuLysGlyAspAspAspThrIleLeuAsn 149
 QY 179 -----AGCAAACTCCGCTCAGATTAAATCT-----GGA 208
 Db 150 AsnTyrLysAlaIleAlaProAspLeuGlnLysLeuAspGlyGlyThrValGlnLeuVala 169
 QY 209 GCGATGCTTTTGTTCGCCGAGGTTCTCAGTGAAGAGACATCTCTGATCTTGGGAAA 268
 Db 170 GlyLeuAspProIleAlaAsnAlaLeuThr---SerThrIleAlaThrAspGlnArgArg 188
 QY 269 AAACAAATGACACAGAGCATTAATCTCAGACGCTGCGCTGGCAATGCTTATGTTTGG 328
 Db 189 MetGluVal-----LeuAlaThrProLeuValAla 198
 QY 329 ATGAGCGTATATATATGATTGAGCGCTGATGCTTGGAGACTGTTCTTGAATCTT 388
 Db 199 ValValaLeuPheLeuValPheGlyGlyValIleAla----- 210
 QY 389 TTGCTATCTGGGACGCTACAGTATTTGGATGCGGCACTGACCTTGCGAGACTCGCT 448
 Db 211 -----AlaCysLeuProValMetValGlyGlyLeuSerIleAlaGlyAlaLeu 226
 QY 449 GGGATT----- 454
 Db 227 GlyIleuArgPheIleAlaLeuPheGlyProValHisPhePheAlaGlnProValVal 246
 QY 455 ---GTTCTTCATAGGGGATGCGCTGATGCAATGTTCTTATTCGAAAGATCCGA 511
 Db 247 SerLeuIleGlyLeuGlyIleAlaValAspTyrGlyLeuPheValValSerArgPheArg 266
 QY 512 GAGGAATTTTATATGCTCAAACTTTAAAAAATCGTGAAGAAA----- 556
 Db 267 GlnGlnIleAlaGluGlyTyrAspThrGluAlaValaValaArgArgThrValMetThrAla 286
 QY 557 GGAATATCAAGGCTTTTGGAGCCATTTTGAATCTTAATCTGACTACAGATATGGCGCTCA 616
 Db 287 GlyArgThrValThrPheSerAlaValLeu-----IleAlaAlaSerGly 301
 QY 617 GCACTTCTTTTCTTCTGATACAGGCGCTATTAAGGCTTGCTTGACATTGATT--- 673
 Db 302 AlaSerLeuLeuLeuLeuProGlnGlyPheValLysSerLeuThrTyrAlaLeuIleAla 321
 QY 674 -----TTAGAATTTTC 685

Db 322 AlaValThrLeuAlaAlaLeuSerIleThrLeuLeuProAlaCysLeuAlaIleLeu 341
Qy 586 TCTTCAATGTTTACGGCTCTT---TTCAAGACTAAATTTTCTTCATGCTGTGGATGAAT 742
Db 342 AlaLysHisValAspAlaLeuGlyValArgThrLeuPheArgValProLeuLeuArgAsn 361
Qy 743 AAGACCAACATACACAGCTTCATATGATGAATAGTTTCGTGGGATAAAGCATGATTC 802
Db 362 TrpArgMetSer-HisAlaCysLeu-----AsnTrp-----LeuAl 373
Qy 803 TTGAGAGGATCAAAAAAATCT-----TTGGCTGTTTCTGGAAGT 841
Db 373 a-AspArgLeuGlnLysThrLysThrArgGluGluValGluAlaGlyPheTrpGlyLysL 393
Qy 842 GTTTTCTTTAGTTGGCTCTCGGTTTGGAGCTCGAATTCGTTTGGGAATG 901
Db 393 euValAsnPheValMetLysArgProLeuValPheAlaIleProIle-ValIleGlyMet 412
Qy 902 GATTTTAAAT-----GGAGGTATGCTTTTACCTTTAAT 934
Db 413 IleLeuLeuValIleProLeuGlyAsnLeuSerPheGlyGlyMetSerGluLysTrpLeu 432
Qy 935 CCAAAAGAGCATGCATCAGCGATGTTCTCAAAATGCGTGCAGAAAGTTGTGCATAAACTA 994
Db 433 ProProAsnAsnAlaValArgGlnSerClnGlnHis-PheAspGlnLeuPheProGlyTy 452
Qy 995 CAGG-----AAGCTGTCTTCTCTAGAGACTTCGCTATTCAAAACATTTGGATCTTCA 1048
Db 452 rArgThrAsnProLeuThrLeuValIleGlnThr-SerAsnHisGlnProValThrAspGl 472
Qy 1049 GAAA-----AGATCAAAATCTATTTAGTGTAGTAAGCTTTAAGCTACTACTAAG 1096
Db 472 nGluIleAlaAspIleArgSerLysAlaMetAlaIleSer-GlyPhe-----487
Qy 1097 CAGATACGAGCTCTCTCTCAAAATTAACGATCATGAGCTGGCTGTTATGTGG-----1149
Db 488 -----IleGluPro-----AspAsnAsnTyrrValAsnMetTrpGlnGluA 501
Qy 1149 -----1149
Db 501 rgThrValAlaProGlyAlaSerLysAspProSerValArgValLeuGlnAsnGlyLeuI 521
Qy 1150 -----GATTGTGTGACAAACAGGCTAGATTTCTCTACGGAACCTCTAAACGA 1198
Db 521 leAsnProAsnAspAlaSerLysLysIleAsn-----GluLeuArgSerIleThrProp 539
Qy 1199 AACGCAAAATTTTGTCTCAAA---GCTAAGCAGCAAACTATCGAAGAAATCGCTTATCAG 1255
Db 539 roLysGlyLeuThrValSerValGlyGlyThrProAlaLeuGluGlnAspSerIle-His 558
Qy 1256 GCGACCATCGGGCTTTTAGGAGCTTTTGGCAATCATCTTGTCTATGTGAGTTTGGCTTT 1315
Db 559 SerLeuValAlaGlnAlaProLeuMetValIleMetLeuIleThrThrMetLeuLeu 578
Qy 1316 GAATGGCAATATGCTTTTACGTGCGGTATGCGCTTAAATCATGACCTTTTGGCTACCTGT 1375
Db 579 MetPhe---LeuAlaPheGlySerPheValLeuProIleLysAlaAlaValMetSerAla 597
Qy 1376 GCAGTCTGTTTATAGCACATTTCTTTTGAAGAAATTCAAATAGAT-----1423
Db 598 LeuThrLeuGlySerThrMetGlyIleLeuThrTrpIlePheValAspGlyHisLeuSer 617
Qy 1424 -----TTGCAAGCATTGTGTCTTTAATGACTGTA 1453
Db 618 LysTrpLeuAsnPheThrProThrProLeuMetValValIleIleAlaLeuValValAla 637
Qy 1454 TTGGGTATTCAATAACAATCTTTCATCATCTTTT-----GATCGATTTCGGAAGAT 1507
Db 638 ValGlyTyrglyLeuAlaThrAspTyrglyValPheLeuValSerArgMetValGluAla 657
Qy 1508 CGCAACGAGCACTGTTTACCCCTATGCTGTTTGTAGTTAATGATGCTCCCTCAAAAGACG 1567
Db 658 ArgAlaGluSerMetSerThrGlnGluAlaValArgIleGly-----671

Qy 1568 TTCAGCCGCGCGTAATGACAAACAGCTACACTCTATCAGTTTGTAAATGCTTTTGT 1627
Db 672 -----ThrAlaSerThrGlyArgLeuIleThrAlaAlaAlaLeuValLeuAlaVal 688
Qy 1628 ATAGCGGCTCTCTCTCTTTAAT-----TTGCAATTTATTATG 1666
Db 689 ValAlaGlySerPheValPheSerAspLeuValMetMetLysTyrrLeuAlaPheGlyLeu 708
Qy 1667 ACCATAGGGATTTCTTCTA---GGAACTTATCGTCTCTTTATATTGCACCACTCTGTG 1723
Db 709 MetAlaAlaLeuLeuLeuAspAlaThrValValArgMetPheLeuValProSerValMet 728
RESULT 40
YHIV_ECOLI
ID YHIV_ECOLI STANDARD; PRT; 1037 AA.
AC P37637;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein yhiv.
GN YHIV OR B3514.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
SEQUENCE FROM N.A.
STRAIN=K12 / MG1655;
RX MEDLINE=94316500; PubMed=8041620;
RA Sofia H.J., Burland V., Daniels D.L., Plunkett G. III, Blattner F.R.;
RT "Analysis of the Escherichia coli genome. V. DNA sequence of the
region from 76.0 to 81.5 minutes."
RL Nucleic Acids Res. 22:2576-2586(1994).
CC -!- FUNCTION: COULD BE A DRUG EFFLUX PUMP (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
(potential).
CC -!- SIMILARITY: BELONGS TO THE ACRB/ACRD/ACRF FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.ebi.ac.uk/ebis-sib.ch/announcement/>
or send an email to license@ebi-sib.ch).
CC EMBL; U00039; AAB18490.1; -;
DR EMBL; AB000427; AAC76539.1; -;
DR EcoGene; EGI2241; Yhiv.
DR InterPro; IPR001036; Acrflvin_res.
DR InterPro; IPR004764; HAE1.
DR Pfam; PF00873; ACR_tran; 1.
DR PRINTS; PRO0702; ACRFLVINRP.
DR TIGRFAMs; TIGR00915; 2A0602; 1.
KW Hypothetical protein; Transmembrane; Inner membrane; Transport;
Complete proteome.
FT DOMAIN 1 9 CYTOPLASMIC (BY SIMILARITY).
FT TRANSMEM 10 28 1 (BY SIMILARITY).
FT DOMAIN 29 339 PERIPLASMIC (BY SIMILARITY).
FT TRANSMEM 340 359 2 (BY SIMILARITY).
FT DOMAIN 360 365 CYTOPLASMIC (BY SIMILARITY).
FT TRANSMEM 366 385 3 (BY SIMILARITY).
FT DOMAIN 386 391 PERIPLASMIC (BY SIMILARITY).
FT TRANSMEM 392 413 4 (BY SIMILARITY).
FT DOMAIN 414 441 CYTOPLASMIC (BY SIMILARITY).
FT TRANSMEM 442 460 5 (BY SIMILARITY).
FT DOMAIN 461 473 PERIPLASMIC (BY SIMILARITY).
FT TRANSMEM 474 496 6 (BY SIMILARITY).
FT DOMAIN 497 536 CYTOPLASMIC (BY SIMILARITY).
FT TRANSMEM 537 555 7 (BY SIMILARITY).
FT DOMAIN 556 871 PERIPLASMIC (BY SIMILARITY).
FT TRANSMEM 871 890 8 (BY SIMILARITY).

FT	DOMAIN	891	896	CYTOPLASMIC (BY SIMILARITY).
FT	TRANSMEM	897	916	9 (BY SIMILARITY).
FT	DOMAIN	917	922	PERIPLASMIC (BY SIMILARITY).
FT	TRANSMEM	923	944	10 (BY SIMILARITY).
FT	DOMAIN	945	972	CYTOPLASMIC (BY SIMILARITY).
FT	TRANSMEM	973	991	11 (BY SIMILARITY).
FT	DOMAIN	992	1004	PERIPLASMIC (BY SIMILARITY).
FT	TRANSMEM	1005	1027	12 (BY SIMILARITY).
FT	DOMAIN	1028	1037	CYTOPLASMIC (BY SIMILARITY).
SQ	SEQUENCE	1037 AA;	111517 MW; 3567502160F51724 CRC64;	
 Alignment Scores:				
Pred. No.:	0.0314	Length:	1037	
Score:	125.50	Matches:	145	
Percent Similarity:	32.43%	Conservative:	107	
Best Local Similarity:	18.54%	Mismatches:	243	
Query Match:	3.79%	Indels:	287	
DB:	1	Gaps:	37	
 US-09-868-987-1 (1-1864) x YHV1_ECOLI (1-1097)				
Oy	113 CCTATTTTAAACGTCCTCAATAAATCATGCCAGTGCTCTCAGGAATTACCACCGT	172		
Dd	285 Proalaalaglyllealliletyleualaiaolaialaleusprlserarg	304		
Oy	173 -----GAAGTAGAACACTGGCCGCACATTAAATCTGGAGCGATGTCTTT	220		
Dd	305 AlaValLySGluGluLeuAsnArgLeuSerAlaTryrheProAlaserylThrVal	324		
Oy	221 GTTCCC-----GAGGTTCTCAGTAAGAAGCATCTTCTGTAT	259		
Dd	325 TyrProTyraSPrlThrThrProPhellegullilerleglnuValPhe-----	341		
Oy	260 CTTGGAAAAACAATGACACAMAGCGATTATCTCACAGTCGTGTGGCTTGCAATGCTT	319		
Dd	342 -----LysThrLeuValGlnAlaLlelleuValPhe-----LeuAlaMetTy	356		
Oy	320 ATTGTTTGATGAGCGGTATATATATGATTGTGAGAGCGCTCATCGCTTGAGAGCTGTCTT	379		
Dd	357 LeuPheLeuGlnAsnPharGlathrLleleleProthrllleAlaValProvalValle	376		
Oy	380 CTGAATCTTTTGGCTTATCTGGGACGCTCTACAGTAATTTGGATGCGCACTCACCTTGTC	439		
Dd	377 LeuGLythrPheAlaLleleuSerAlaValglYethrLleAsnthrLeuthrMetPhe	396		
Oy	440 GGACTCGCTGGGATGCTCTTGCTATGAGGAGCGCCATGATGAATGTTCTGTATTC	499		
Dd	397 GlyMet-----ValLeuAlalleGlyLeuLeuValAspaRaLaileValVal	413		
Oy	500 GAAAGATCCGAGAGGAATTTTATTTGCTCAAGCTCT-----AAAAATCTGTAGA	553		
Dd	414 GluAsnVal---GluArgValllelaInuBrylsueuProprolysguilaThnHis	432		
Oy	554 AAGAGTATATCCAGACCTTTGGAGCCATTTTGTGATCTTAACCTGACAGTATTTGGCC	613		
Dd	433 LysSerMetGlyGlnleGlnArgAlaLeuVal-----GlylleAlaValleuSer	450		
Oy	614 TCAGCACTT-----CTTTCTCTCTAAT-----ACAGGCGCTATT-----AAAGGGT	658		
Dd	451 AlaValPheMetProMetAlaPheMetserGlyAlaInuGlyGuilleTyraGlnPhe	470		
Oy	659 GCATTGACATTAATTTTGGAAATTTTCTCTTCATGTTTACGGCTTTTCATGACTAAA	718		
Dd	471 SerlleThrLeulleSerSerMetLeuSerValPheValAlaMetSerleuthrPro	490		
Oy	719 TTT-----	721		
Dd	491 AlaLeuCysAlaThrLleLeuYSaAlaAProglnGlyGlnHslYsProAsnAlaLeu	510		
Oy	722 -----TTCTCATGCTGTGGTGAATAGAACCCAAACAT-----	754		
Dd	511 PheAlaArgPheAsnthrLeuPheGluYsSerThrGlnHslTyrrThrAspSerThrArg	530		

QY	755	ACAAAGTTGGATATGATGAATAAGTTCCG	-----GGGATA	790
		:::	::::::::::	
Db	531	SerLeuLeuArgCysThrGlyArgTyrMetValValTyrLeuLeuIleCysAlaGlyMet	-----	550
QY	791	AAGCATGATTTCTTGTAGA	-----	808
Db	551	AlaValLeuPheLeuArgThrProThrSerPheLeuProGluGluAspGlnGlyValPhe	-----	570
QY	809	-----	GGATCCAAAAA	826
Db	571	MetThrThrAlaGlnLeuProSerGlyAlaThrMetValAenThrThrValValLeuGln	-----	590
QY	827	GCCTGTTCCGGA	-----AGGTTTTCTTTA	853
		:::		
Db	591	GlnAlaThrAspTyrTyrLeuThrTyrGluTyrAspMetValGlnSerValPheThrVal	-----	610
QY	854	GGTTGCGTTGCTCTCGGTTTGA	-----	877
Db	611	Gly-----	GlyPheGlyPheSerGlyGlnGlyGlnAsnGlyLeuAlaPhe	626
QY	878	-----GCCTGGAATTCGTTTGGGATGCAT	-----	904
		:::		
Db	627	IleSerLeuLysProThrPergGluArgValGlyGluGlnAsnSerValThrAlaIleIle	-----	646
QY	905	-----TTTAAGAGAGGTATGCCTTACCTTTAAT	-----	934
Db	647	GlnArgAlaMetIleAlaLeuSerSerIleAsnLysAlaValValPheProPheAsnLeu	-----	666
QY	934	-----	-----	934
Db	667	ProAlaValAlaGluLeuGlyThrAlaSerGlyPheAspMetGluLeuLeuAspAsnGly	-----	686
QY	934	-----	-----	934
Db	687	AsnLeuGlyHisGluLysLeuThrGlnAlaArgAsnGluLeuLeuSerLeuAlaAlaGln	-----	706
QY	935	-----CCAAAGAGCATGCGCATCAGGAGATGTTGCTCAAAATCGT	-----	973
Db	707	SerProAsnGlnValThrGlyValAlaArgProAsnGlyLeuGluAspThrProMetPheLys	-----	726
QY	974	GGCAAAAGTT---GNGCATAACTACAGAGAGCTGGTCTTCTTCACAGACTTCGCT	-----	1027
Db	727	ValasnValasnAlaAlaLysAlaGluAlaMetGlyValAlaLeuSerAspIleAsnGln	-----	746
QY	1028	---ATTCAAACA---TTTGGATCTTCAGAAAAGATCAAAATCTATTTTAGTGAT	-----	1075
Db	747	ThrIleSerThrAlaPheGlySerSer-----TyrValAsnAspPheLeu	-----	761
QY	1076	-----	AAAGCTTTAGC	1087
Db	762	AsnGlnGlyArgValLysLysValTyrValGlnAlaGlyThrProPheArgMetLeuPro	-----	781
QY	1088	-----TATTAAGACGATACAGAGCTCTCTCTCTAAATTAACATC	-----	1129
		:::		
Db	782	AspAsnIleAsnGlnTPrpTyrValaArgAsnAlaSerGlyThrMetAlaProLeuSerAla	-----	801
QY	1130	ATGAGC-----TGGCGTTATTTGGGATGTTGTTCAGAAACAGGCTCAGATT	-----	1177
Db	802	TyrSerSerThrGluTPrpThrTyr-----GlySerProArgLeuGlu	-----	815
QY	1178	-----CTTACGAGAACTCTTAA	-----	1195
Db	816	ArgTyrAsnGlyIleProSerMetGluIleLeuGlyGluAlaAlaAlaGlyLysSerThr	-----	835
QY	1196	GGAAACGCA---AAATTTGGTCAAAAGGTACAGCAGCAACTATCCAAAGAAATGCGTTAT	-----	1252
		:::		
Db	836	GlyAspAlaMetLysPheMetAlaAspLeuValAlaLysLeuProAlaGlyValGlyLys	-----	855
QY	1253	CAGGCGACCATCGGCGTT-----TTA	-----	1273
Db	856	SerTPrpThr---GlyLeuSerTyrGlnGluAlaLeuSerSerAsnGlnAlaProAlaLeu	-----	874
QY	1274	GGAGCTTTGGCAATCATCTTGCTCTATGATGACGTTTCGCTT-----GAATGCGCA	-----	1324


```

Db 875 TyrAlaLeuSerLeuValValPheLeuAlaLeuAlaLeuTyrGluSerTrpSer 894
Oy 1325 TATGCTTTCAGTCCGTA-----TGCCTTTAAATTCATGACCTTTTGGCTACC 1372
Db 895 IleProPheSerValMetLeuValProLeuGlyValValGlyAlaLeuLeuAlaThr 914
Oy 1373 TGTGCGCTCTGTTTATAGCACATCTCTTTTGAAGAAATTCAAATAGATTGCAAGCC 1432
Db 915 -----AspLeuArgGlyLeuSerAsnAspValTyrPhe 925
Oy 1433 ATTGGTCTTTAATGACTGATTGGGATTCATTAAACAATCTTTGATCATTTTGGAT 1492
Db 926 GlnValGlyLeuThrThrIleGlyLeuSerAlaValAsnAlaIleLeuValGlu 945
Oy 1493 CGTATTCGTGAAGATGCCAAGCAACCTGTTTACCCCTATGCATCTTTTAAATGAT 1552
Db 946 PheAlaValGluMetGlnLysGluGlyLysThrProIleGluAlaIleGluAla 965
Oy 1553 GCCCTTCAAAAGAGCTTCAGCCGACGGATGACACAGCTACACTCTATCATGTTTGT 1612
Db 966 AlaArgMetArgLeuArgProIleLeuMetThrSerLeuAlaPheIleLeuGlyValLeu 985
Oy 1613 TTAATGCTTTGTTTATAGGC---GGCTCCTCTGCTTTTAAATTTTGCATTTTATGACC 1669
Db 986 ProLeuValIleSerHisGlyAlaGlySerGlyAlaGlnAsn----- 999
Oy 1670 ATAGGGATTTCTTACGAACTTATCGCTCTCTCTATATTCACCACTCTGTTGTTT 1729
Db 1000 --AlaValGlyThrGlyValMetGlyGlyMetPheAlaAlaThrValLeuAlaIleTyr 1018
Oy 1730 ATGCTC 1735
Db 1019 PheVal 1020

RESULT 41
MMLB MYCLE
ID MMLB MYCLE STANDARD; PRT; 1014 AA.
AC 006079;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Putative membrane protein mmp111.
GN MML11 OR M2617 OR MLC622.16C.
OS Mycobacterium leprae.
OC Bacteria; Actinobacteria; Actinobacteridia (class); Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=TN;
RX MEDLINE=21128732; PubMed=11234002;
RA Cole S.T., Biglmeier K., Parkhill J., James K.D., Thomson N.R.,
RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA Barrell B.G.;
RT "Massive gene decay in the leprosy bacillus.";
RL Nature 409:1007-1011(2001).
CC CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC CC -1- SIMILARITY: BELONGS TO THE MMLP FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```

CC EMBL; Z95398; CAB08803.1; -.
DR EMBL; AL583926; CAC32149.1; ALT_INIT.
DR Leproma; ML2617; -.
DR InterPro; IPR000731; HMGCR/patch_5TM.
DR InterPro; IPR004869; MMLP.
DR Pfam; PF031176; MMLP; 1.
DR PROSITE; PSS0156; SSD; 1.
KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 13 33
FT TRANSMEM 188 208 POTENTIAL.
FT TRANSMEM 214 234 POTENTIAL.
FT TRANSMEM 235 255 POTENTIAL.
FT TRANSMEM 279 299 POTENTIAL.
FT TRANSMEM 311 331 POTENTIAL.
FT TRANSMEM 373 393 POTENTIAL.
FT TRANSMEM 409 429 POTENTIAL.
FT TRANSMEM 530 550 POTENTIAL.
FT TRANSMEM 560 580 POTENTIAL.
FT TRANSMEM 598 618 POTENTIAL.
FT TRANSMEM 649 669 POTENTIAL.
FT TRANSMEM 671 691 POTENTIAL.
SQ SEQUENCE 1014 AA; 109875 MW; 5BB0D797AFF58D79 CRC64;

Alignment Scores:
Pred. No.: 0.0256 Length: 1014
Score: 124.50 Matches: 103
Percent Similarity: 33.22% Conservative: 85
Best Local Similarity: 18.20% Mismatches: 203
Query Match: 3.76% Indels: 175
DB: 1 Gaps: 24

US-09-868-987-1 (1-1864) x MMLB MYCLE (1-1014)
Oy 374 GTTCTCTGAATCTTTTCTTCTGCGGAGCTCTACAGTATTTGGATGCGCCACTCACC 433
Db 189 IleLeuIleValLeuLeuAlaValPheGlySerLeuAlaAlaAlaValProLeuAla 208
Oy 434 TTGTCAGGACTCGCTGGGATTTGTTCTTCTGATGGG----- 469
Db 209 Leu---GlyValCysThrValValValThrMetGlyLeuValAspLeuValSerMetHis 227
Oy 470 -----ATGCCCGTA 478
Db 228 ThrIleMetSerValPheValThrSerThrValSerMetPheGlyIleAlaLeuAlaVal 247
Oy 479 GATGCAATGTTCTTGATTTCGAAGAATCCGAGAGCAATTTTATTTGCTCAAGCTTT 538
Db 248 AspTyrSerLeuPheIleLeuMetArgPheArgGluLeuArgSerGlyArgGlnPro 267
Oy 539 AAAAAATCTGTAGAAAAAGGATATATACCAAGGCTTTTGGAGCCATTTTGTATTCTAACTTG 598
Db 268 GlnGluAlaValAspAlaAlaMetAlaThrSerGlyLeuAlaValValLeuSerGlyMet 287
Oy 599 ACTACAGTATTGGCTCAGCACTCTTTTCTTCTAGATACAGGGCCTATTAAAGGTTT 658
Db 288 ThrValIleAlaSerLeuThrGlyIleTyrLeuIleAsnThrAlaAlaLeuLysSerMet 307
Oy 659 GCTTTGACATGATTTTAGGAATTTCTCTCAATGTTTACGGCTCTTTTCATGACT--- 715
Db 308 AlaThrGlyAlaIleLeuAlaValAlaIleAlaMetLeuAlaSerIleThrLeuThrPro 327
Oy 716 -----AAATTTTCTTCATGCTCTGGATG 739
Db 328 AlaAlaLeuAlaThrPheGlyArgAlaAlaValLysArgSerValLeuMetHisTrpSer 347
Oy 740 AATAAGACCAACATACACAGATTGTCATATGATGAATTAAGTTCTGGGGG---ATAAAGCAT 796
Db 348 GlnArgSerGluCysThrGlnSerLeuPheThrThrArgTrpValGlyTrpValMetHis 367
Oy 797 GATTTCTTGAGAGATGCCAAAAACTTTGG-----GCTGTTTCTGGAAGTGT 844
Db 368 -----ArgProTrpIleSerAlaSerAlaAlaSerThrIleLeu 380

```


FT TRANSMEM 852 871 8 (BY SIMILARITY).
 FT DOMAIN 872 877 CYTOPLASMIC (BY SIMILARITY).
 FT TRANSMEM 878 897 9 (BY SIMILARITY).
 FT TRANSMEM 898 903 PERIPLASMIC (BY SIMILARITY).
 FT TRANSMEM 904 925 10 (BY SIMILARITY).
 FT TRANSMEM 926 953 CYTOPLASMIC (BY SIMILARITY).
 FT TRANSMEM 954 972 11 (BY SIMILARITY).
 FT TRANSMEM 973 985 PERIPLASMIC (BY SIMILARITY).
 FT TRANSMEM 986 1008 12 (BY SIMILARITY).
 FT DOMAIN 1009 1025 CYTOPLASMIC (BY SIMILARITY).
 SQ SEQUENCE 1025 AA; 111010 MW; EF00BB4E7B301008 CRC64;

Alignment Scores:
 Pred. No.: 0.0309 Length: 1025
 Score: 123.50 Matches: 130
 Percent Similarity: 33.33% Conservative: 105
 Best Local Similarity: 18.44% Mismatches: 243
 Query Match: 3.73% Indels: 227
 DB: 1 Gaps: 28

US-09-868-987-1 (1-1864) x YEGO_ECOLI (1-1025)

Qy 245 ACGATCTCTTCTGATCTTGGGAAAAACAATGTACACAAGGCATTATCTCAGCATGCTGT 304
 Db 326 ThrileArgAlaSerLeuGluGluValGluGluThrLeu---IleIleSer----- 341

Qy 305 GGCTTGGCAATGCTTATGTTTGGATGAGCGTATATATATAGATTGGA-----GGCGTC 358
 Db 342 ---ValAlaLeuValIleLeuValValPheLeuPheLeuArgSerGlyArgAlaThrIle 360

Qy 359 ATCCCTTCGGAGCTGCTCTCTGATCTTCTGATCTTCTGATCTTCTGATCTGCGGATGCGGTA 478
 Db 361 IleProAlaValSerValProValSerLeuIleGlyThrPheAlaAlaMetTyrLeuCys 380

Qy- 419 GATCGGCACCTCACGTCTGAGGATTCGCGGATTTGTTGCTTATGCGGATGCGGTA 478
 Db 381 GlyPheSerLeuAsnLeuSerLeuMetAlaLeuThrIleAlaThrGlyPheValVal 400

Qy 479 GATCAAAATGCTTGTATGATTCGAAAGATCCGAGAGGAATTT-----TTA 523
 Db 401 AspAlaIleValValLeuGluAsnIleAlaArgHisLeuGluAlaGlyMetLysPro 420

Qy 524 TTGTCTCAAGCTCTTAAAAATCTGTAGAAAAGATATACCAAGGCTTTTGGAGCCATT 583
 Db 421 LeuGlnAlaLeuGlnGlyThrArgGluValGlyPheThr-----Val 435

Qy 584 TTTGATTCTAACTTGACATGATGTCCTGCGCTCAGCACTTCTTTCTCTCTAGATACAGG 643
 Db 436 LeuSerMetSerLeuSerLeuValAlaValPheLeuProLeuLeuMetGlyGlyLeu 455

Qy 644 CCT-----ATTAAAGGCTTGTGCTTGTGATGTTTGGAAATTTCTCTTCAATG 694
 Db 456 ProGlyArgLeuLeuArgGluPheAlaValThrLeuSerValAlaIleGlyLeSerLeu 475

Qy 695 TTTACGGCTCTTTTCATGACTAAATTTTCTTCATGCTGTGATGATTAAGACC----- 748
 Db 476 LeuValSerLeuThrLeuThrPro---MetMetCysGlyTrpMetLeuLeuAlaSerLys 494

Qy 749 -----CAACATACACAGTTCATATGATGAATAAG---TTCGTGGGATTAAGCATCAT 799
 Db 495 ProArgGluGlnLysArgLeuArgGlyPheGlyArgMetLeuValAlaLeuGlnGlnGly 514

Qy 800 TTCTTGAGAGGATGCAAAAACTTTGGCGCTGTTCTGGAAGT-----GTTTTT 847
 Db 515 TyrGlyLysSerLeuLys-----TrpValLeuAsnHisThrArgLeuValGlyValVal 532

Qy 848 CTTTGTAGTTCGCTGCTCGGTTTGGAGCGCTGGAATTCGGTT----- 892
 Db 533 LeuLeuGlyThrIleAlaLeuAsnIleTrpLeuTyrIleSerIleProLysThrPhePhe 552

Qy 893 -----TTGGGA----- 898
 Db 553 ProGluGlnAspThrGlyValLeuMetGlyGlyIleGlnAlaAspGlnSerIleSerPhe 572

Qy 899 -----ATCGATTTT----- 907
 Db 573 GlnAlaMetArgGlyLysLeuGlnAspPheMetLysIleIleArgAspAspProAlaVal 592

Qy 908 -----AAAGGAGGGTATGCTTACC 928
 Db 593 AspAsnValThrGlyPheThrGlySerArgValAsnSerGlyMetMetPheIleThr 612

Qy 929 TTTAATCAAAAGAGCATGCGCATGCGATGCTCTCAAAATGCGTGGCAAGTTGTGTCAT 988
 Db 613 LeuLysProArgAspGluArgSerGluThrAlaGlnGlnIleIleAspArgLeuArgVal 632

Qy 989 AACTA---CAGGAAGCTGGT-----CTTCTCTAGAGATTCGCTATT 1030
 Db 633 LysLeuAlaLysGluProGlyAlaAsnLeuPheLeuMetAlaValGlnAspIleArgVal 652

Qy 1031 -----CAAAACATTGGCATCTTCAGAAAAGATCAAAATCTATTTTAGTGAAGACCTTTA 1084
 Db 653 GlyGlyArgGlnSerAsnAlaSerTyrGlnTyrThrLeuLeuSerAspAspLeuAla 672

Qy 1085 -----AGCTATATAAGCAGATACGAGCCTCTCTCTAAAATTAACGATCATG----- 1132
 Db 673 LeuArgGluTrpGluProLysIleArgLysLysLeuAlaThrLeuProGluLeuAlaAsp 692

Qy 1133 -----AGC 1135
 Db 693 ValAsnSerAspGlnGlnAspAsnGlyAlaGluMetAsnLeuValTyrAspArgAspThr 712

Qy 1136 TGGCGTTATTTGCGGATTTGTCAGAAAAGCGCTAGATTTCTCTACGGAACTCTAAA 1195
 Db 713 MetAlaArgLeuGlyIleAspValGln-----AlaAlaAsnSerLeu 726

Qy 1196 CGAAACCAAAATTTTGGTCAAAGCTAAGCAGCAAACTATCGAAAGAAATG----- 1246
 Db 727 LeuAsnAsnAlaPheGlyGlnArgGlnIleSerThrIleTyrGlnProMetAsnGlnTyr 746

Qy 1247 -----CGTTATCAGCGCACCATCGGCTTTTAGGAGCTTTG 1282
 Db 747 LysValValMetGluValAspProArgTyrThrGlnAspIleSerAlaLeuGluLysMet 766

Qy 1283 GCATCATC-----TTGCTCTATGTAGTTGCGCTTTGATGCAATAT 1327
 Db 767 PheValIleAsnAsnGluGlyLysAlaIleProLeuSerTyrPheAlaLysTrpGlnPro 786

Qy 1328 GCT----- 1330
 Db 787 AlaAsnAlaProLeuSerValAsnHisGlnGlyLeuSerAlaAlaSerThrIleSerPhe 806

Qy 1330 ----- 1330
 Db 807 AsnLeuProThrGlyLysSerLeuSerAspAlaSerAlaAlaIleAspArgAlaMetThr 826

Qy 1331 -----TTAGTCCCGTATGCGCTTTTAATTCAT 1357
 Db 827 GlnLeuGlyValProSerThrValArgGlySerPheAlaGlyThrAlaGlnValPheGln 846

Qy 1358 GACCTTTTGGCTACCTGCGAGTCTCTTTATAGCACATTTCTTTTGAAGAAAATCAA 1417
 Db 847 GluThrMetAsnSerGlnValIleLeuIleAlaAlaIleAlaThrValIleVal 866

Qy 1418 ATAGATTG----- 1426
 Db 867 LeuGlyIleLeuTyrGluSerTyrValHisProLeuThrIleLeuSerThrLeuProSer 886

Qy 1427 CAAGCCATTGGCTTTA----- 1444
 Db 887 AlaGlyValGlyAlaLeuLeuAlaLeuLeuPheAsnAlaProPheSerLeuIleAla 906

Qy 1445 -----ATGACTGTATGGGATTCATTAACATATCTTTGATCATTTTGCAT 1492
 Db 907 LeuIleGlyIleMetLeuLeuIleGlyIleValLysLysAsnAlaIleMetMetValAsp 926

```

QY 1493 CGATTTCGTAAGATCCGCAAGCAACCTGTTTACCCCTATGCATGTTTATGATATGAT 1552
DB 927 PhealaleuGluAlaGlnAghIaGlyAsnLeuThrProGlnGluAlaIlePheGlnAla 946
QY 1553 GCCCTTCAAAAGAGCTTCAGCCGACGCGTACCAACAGCTTACACTTCAGTTTGG 1612
DB 947 CysLeuLeuArg---PheArgProIleMetMetThrThrLeuAlaIaIaLeuPheGlyAla 965
QY 1613 TTATATGCTTTTGTATTATAGGCGG---TCCTCTGCTCTTTAATTTGATTTATATGACC 1669
DB 966 LeuProLeuValIleuSerGlyGlyAspGlySerGluLeuArgGlnProLeuGlyIleThr 985
QY 1670 ATA-----GGGATTCCTTCAGGAACTTATCGCTCTTTATATGACCACTCTGTGTG 1723
DB 986 IleValGlyGlyLeuValMetSerGlnLeuThrLeuThrThrProValValTyr 1005
QY 1724 TTGTTTATGCTCCGT 1738
DB 1006 LeuPhePheAspArg 1010

RESULT 43
MEXB_PSEAE STANDARD; PRT; 1046 AA.
ID MEXB_PSEAE
AC PE2002;
DT 01-OCT-1996 (Rel. 34, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Multidrug resistance protein mexb (Multidrug-efflux transporter mexb).
GN MEXB OR PA0426.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CD10;
RX MEDLINE=95058196; PubMed=7968531.
RA Poole K., Heinrichs D.E., Neshat S.;
RT "Cloning and sequence analysis of an EnvC homologue in Pseudomonas
aeruginosa: regulation by iron and possible involvement in the
secretion of the siderophore pyoverdine.";
RT Mol. Microbiol. 10:529-544(1993).
RL [2]
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=2043737; PubMed=10984043;
RA Stover C.K., Pham X.-O.T., Erwin A.L., Mizoguchi S.D., Warrenner P.,
Hickey M.J., Brinkman F.S.L., Hutnagle W.O., Kowalik D.J., Lagrou M.,
Garber R.L., Goltz L., Tolentino E., Westbrock-Madman S., Yuan Y.,
Brody L.L., Coulter S.N., Folger K.R., Kas A., Ladisig K., Lim R.M.,
Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
opportunistic pathogen.";
RT Nature 406:959-964(2000).
RL [3]
RN [3]
RP TOBOLGY.
RX MEDLINE=99214619; PubMed=10187844;
RA Guan L., Ehrmann M., Yoneyama H., Nakae T.;
RT "Membrane topology of the xenobiotic-exporting subunit, MexB, of the
MexA/B-Operm extrusion pump in Pseudomonas aeruginosa.";
RT J. Biol. Chem. 274:10517-10522(1999).
CC -1- FUNCTION: XENOBIOTIC-EXPORTING SUBUNIT OF THE MEXA-MEXB-OPRM
EFFLUX SYSTEM THAT CONFERS MULTIDRUG RESISTANCE. IMPLICATED IN THE
SECRETION OF THE SIDEROPHORE PYOVERDINE.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane.
CC -1- SUBCELLULAR LOCATION: BY GROWTH UNDER SEVERE IRON LIMITATION.
CC -1- SIMILARITY: BELONGS TO THE ACMB/ACRD/ACRF FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way

```

```

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L1616; AAA74437.1; -.
DR EMBL; AE004479; AAG03815.1; -.
DR InterPro: IPR001036; Acriflavin_res.
DR InterPro: IPR004764; HAEI.
DR Pfam; PF00873; ACR_tran; 1.
DR PRINTS; PR00702; ACRIFLAVINRP.
DR TIGRPFAM; TIGR00915; 2A0602; 1.
KM Transmembrane; Inner membrane; Transport; Antibiotic resistance;
KM Complete proteome.
FT DOMAIN 1 9 CYTOPLASMIC.
FT TRANSMEM 10 28 PERIPLASMIC.
FT DOMAIN 29 339 PERIPLASMIC.
FT TRANSMEM 340 359 CYTOPLASMIC.
FT DOMAIN 360 365 CYTOPLASMIC.
FT TRANSMEM 366 385 PERIPLASMIC.
FT DOMAIN 386 391 PERIPLASMIC.
FT TRANSMEM 392 413 CYTOPLASMIC.
FT DOMAIN 414 441 CYTOPLASMIC.
FT TRANSMEM 442 460 PERIPLASMIC.
FT DOMAIN 461 473 PERIPLASMIC.
FT TRANSMEM 474 496 CYTOPLASMIC.
FT DOMAIN 497 538 CYTOPLASMIC.
FT TRANSMEM 539 557 PERIPLASMIC.
FT DOMAIN 558 571 PERIPLASMIC.
FT TRANSMEM 572 891 CYTOPLASMIC.
FT DOMAIN 892 897 PERIPLASMIC.
FT TRANSMEM 898 917 PERIPLASMIC.
FT DOMAIN 918 923 PERIPLASMIC.
FT TRANSMEM 924 945 CYTOPLASMIC.
FT DOMAIN 946 972 CYTOPLASMIC.
FT TRANSMEM 973 991 PERIPLASMIC.
FT DOMAIN 992 1004 PERIPLASMIC.
FT TRANSMEM 1005 1027 CYTOPLASMIC.
FT DOMAIN 1028 1046 CYTOPLASMIC.
FT CONFLICT 90 90 I -> T (IN REF. 1).
SQ SEQUENCE 1046 AA; 112790 MW; PF918553868ED2 CRC64;

Alignment Scores:
Pred. No.: 0.031 Length: 1046
Score: 123.50 Matches: 146
Percent Similarity: 30.70% Conservative: 101
Best Local Similarity: 17.83% Mismatches: 249
Query Match: 3.73% Indels: 295
DB: 1 Gaps: 32

US-09-868-987-1 (1-1864) x MEXB_PSEAE (1-1046)
QY 146 AGTGTCTCAGGAAATTTTATCCCACTGGAAGTGCAAACTCGCTTCAATTTAAATCT 205
DB 276 SerIleAsnAlaGlnPheAsnGlySerProAlaSerGlyIleAlaIleLysLeuAlaThr 295
QY 206 GGACCGATGCTTTGTTGTCCTCCAGGTTCTCAGTGAAGAAGCATCTCTCT 256
DB 296 GlyAlaAsnAlaLeuAspThrAlaLysAlaIleArgGlnThrIleAlaAsnLeuGluPro 315
QY 257 -----GATCTTGGAAAAAACAATGT-----ACACAAGCATTTATCTCAGCA 298
DB 316 PheMetProGlnGlyMetLysValValIleValIleProTyrAspThrThrProValIleSerAla 335
QY 299 -----TCGTGGCTTGGCAATGCTTATGTTGTTGATGAGCGTA 337
DB 336 SerIleHisGluValValIleThrLeuGlyGlyAlaIleLeuLeuValPheLeuValMet 355
QY 338 TATTATAGATTGGAGGC-----GTCATCGCTTGGAGCGTGT 376
DB 356 TyrLeuPheLeuGlnAsnPheArgAlaThrLeuIleProThrIleAlaValProValVal 375
QY 377 CTTCTGAATCTTTTCTTATCTGAGCGAGCTCTACAGTATTTGATGGCCACTGACCTTG 436

```

```

Db 376 LeuLeuGlyThrPheGlyValLeuAlaPheGlyPheSerIleAsnThrLeuThrMet 395
Qy 437 TCAGAGCTCGTGGGATGTTCTGCTATGGGARGCGGTAGATGCAAAATGTTCTGTA 496
Db 396 PheGlyMet-----ValLeuAlaIleGlyLeuLeuValAspAlaIleValVal 412
Qy 497 TTCGAAAGATCCGAGAGGAATTTTATTGCTCAAGTCTT-----AAAAATCTGTA 550
Db 413 ValGluAsnVal-----GluArgValMetAlaGluGlyLeuSerProArgGluAlaAla 431
Qy 551 GAAAAGGATATACCAAGGCTTTTGGAGCCATTTTGTGATCTTAACCTGACTACAGTATTG 610
Db 432 ArgYysSerMetGlyGlnIleGlnGlyAlaLeuVal-----GlyIleAlaMetValLeu 449
Qy 611 GCCTCAGCAGCTTCTT-----TTCCTCCTAGATACAGGCGCTATT-----AAAGGG 655
Db 450 SerAlaValPheLeuProMetAlaPhePheGlyGlySerThrGlyValIleIleArgGln 469
Qy 656 TTGCTTTGACATTGATTAGGAATTTCTCTCAATGTTTACGGCTCTTTTCATGACT 715
Db 470 PheSerIleThrIleValSerAlaMetAlaLeuSerValIleValAlaLeuIleLeuThr 489
Qy 716 -----AAA 718
Db 490 ProAlaLeuCysAlaThrMetLeuLysProIleGluLysGlyAspHisGlyGluHisLys 509
Qy 719 TTTTCTTCATGCTGTGATGAATAAGAACCAACATACACAGTTGCATATGATGAATAAG 778
Db 510 GlyGlyPhePheGlyTrpPheAsnArg-----Met 519
Qy 779 TTCGTGGGATAAGCAGTATTTCTTGAGAGATGCAAAAACTTTGG-----826
Db 520 PheLeuSerThrThrHisGlyTyrrGluArgGlyValAlaSerIleLeuLysHisArgAla 539
Qy 827 -----GCTGTTCTCGAAGTGTTTCTTTTAGTTGGCTT 862
Db 540 ProTyrrLeuLeuIleTyrrValIleValAlaGlyMetIleTrpMetPheThrArgIle 559
Qy 863 GCTCTCGGGTTT-----874
Db 560 ProThrAlaPheLeuProAspGluAspGlnGlyValLeuPheAlaGlnValGlnThrPro 579
Qy 874 -----874
Db 580 ProGlySerSerAlaGluArgThrGlnValValAspSerMetArgGluTyrrLeuLeu 599
Qy 874 -----874
Db 600 GluLysGluSerSerSerValSerSerValPheThrValThrGlyPheAsnPheAlaGly 619
Qy 875 -----GGA 877
Db 620 ArgGlyGlnSerSerGlyMetAlaPheIleMetLeuLysProTrpGluGluArgProGly 639
Qy 878 GCCTGGAATTCGGTTTGGGAATG-----GATTTTAAA 910
Db 640 GlyGluAsnSerValPheGluLeuAlaLysArgAlaGlnMetHisPhePheSerPheLys 659
Qy 911 GGAGGGATGCTTTTACCTTTTAATCCA-----937
Db 660 AspAlaMetValPheAlaPheAlaProProSerValLeuGluLeuGlyAsnAlaThrGly 679
Qy 937 -----937
Db 680 PheAspLeuPheLeuGlnAspGlnAlaGlyValGlyHisGluValLeuGluAlaArg 699
Qy 938 -----AAAGCAGATGGC 949
Db 700 AsnLysPheLeuMetLeuAlaAlaGlnAsnProAlaLeuGlnArgValArgProAsnGly 719
Qy 950 ATCAGCGATGTTGCTCAATCGGTGGCAAGTT---GTGCATAAACTACAGGAGCTGGT 1006
Db 720 MetSerGluProGlnTyrrLysLeuGluIleAspAspGluLysAlaSerAlaLeuGly 739

```

```

Qy 1007 CTTTCTTCTAGAGACTTCCGGT-----ATTCAA 1033
Db 740 ValSerLeuAlaAspIleAsnSerThrValSerIleAlaTrpGlySerSerTyrrValAsn 759
Qy 1034 ACATTTCGATCTTCAGAAAGATCACA---ATCTATTATT-----1069
Db 760 AspPheIleAspArgGlyArgValTyrrValArgValTyrrLeuGlnGlyArgProAspAlaArg 779
Qy 1070 -----AGTGATAAGCTTTAAGCTATAGCTACTAGACAGATACGAGCTCTCTCTCTAAA 1120
Db 780 MetAsnProAspAspLeuSerLysTrpTyrrValArgAsnAspLysGlyGluMetValPro 799
Qy 1121 TTAACGATCATGAGC-----TGG-----CGTTATTGT 1147
Db 800 PheAsnAlaPheAlaThrGlyLysTrpGluTyrrGlySerProLysLeuGluArgTyrrAsn 819
Qy 1148 GGG-----ATTGTTGTCAGAAACAGGCTAGATTCTCTACGGAACTCT 1192
Db 820 GlyValProAlaMetGluIleLeuGlyGluProAlaProGlyLeuSerSerGlyAspAla 839
Qy 1193 AAACGAAACGCAAAATTTTGTCAAGTAAGCAGCAAACTATCGAAGAAATGCGTTAT 1252
Db 840 MetAlaIleValGlu-----GluIleValLysGlnLeuProLysGlyValGlyTyrr 856
Qy 1253 CAGCGGACCATCGCGTT-----1270
Db 857 SerTrpThr---GlyLeuSerTyrrGluGluArgLeuSerGlySerGlnAlaProAlaLeu 875
Qy 1271 -----TTAGAGCTTTGGCAATCATCTTCTCTATGTAGTTTGGCTTTGATGCA 1324
Db 876 TyrrAlaLeuSerLeuLeuValValPheLeuCysLeuAlaAlaLeuTyrrGluSerTrpSer 895
Qy 1325 TATGCTTTTCACTGCGCTA-----TGCCTTTAATTATCATGACCTTTTGGCTACC 1372
Db 896 IleProPheSerValMetLeuValValProLeuGlyValIleGlyAlaLeuLeuAlaThr 915
Qy 1373 TGTCACTCTTGTTTATAGCACATTTCTTTTGAAGAAAAATTCAAATAGATTTCGAAGCC 1432
Db 916 Ser-----MetArgGlyLeuSerAsnAspValPhePhe 926
Qy 1433 ATTGGTCTTAAATGACTGTATGGGGTATTCATTAAACAATACTTTTGCATCTTTTGCAT 1492
Db 927 GlnValGlyLeuLeuThrThrIleGlyLeuSerAlaLysAsnAlaIleLeuIleValGlu 946
Qy 1493 CGTATTCTGAAGATCGCAAGCAACCTGTTTACCCTATGCATGTTTGTAGTTAATCAT 1552
Db 947 PheAlaLysGlu-----LeuHisGluGlnGlyLysGlyIleValGluAla 961
Qy 1553 GCCCTTCAAAAGACGTTTCAGCCGC-----ACGGTAATGACAACA---GCTACAACT 1600
Db 962 AlaIleGluAlaCysArgMetArgLeuArgProIleValMetThrSerLeuAlaPheIle 981
Qy 1601 CTATCAGTTTGTTAATGCTTTTGTATAGCGGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1660
Db 982 LeuGlyValValProLeuAlaIleSerThrGlyAlaGlySerGlySerGlnHisAla---1000
Qy 1661 ATTATGACCATAGGATTTCTTAGGAACCTTTATCGCTCTTTTATATTGACCACTCTG 1720
Db 1001 -----IleGlyThrGlyValIleGlyMetValThrAlaThrValLeu 1015
Qy 1721 TTGTTGTTTATGGTC 1735
Db 1016 AlaIlePheTrpVal 1020

```

```

RESULT 44
MMLE MYCTU          STANDARD;      PRT;    966 AA.
ID   MMLE MYCTU
AC   OS3653;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Putative membrane protein mmpL11.

```

GN MMBP1 OR RV0202C OR MT0212 OR MT033.10C.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 NCBI_TaxID=1773;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=H37RV; PubMed=9634230;
 RX MEDLINE=98295987; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eiglmeier K., Gae S., Barry C.E. III, Tekala F.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holroyd S.,
 RA Hornsby T., Jagers K., Krogh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
 RA Sultson J.E., Taylor K., Whitehead S., Barrell B.G.;
 RA "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence."
 RL Nature 393:537-544(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CDC 1551 / Oshkosh;
 RA Fleisemann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 RA Peterson J., Deboy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
 RA Kolony J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
 RA Delcher A., Utterback T., Weidman J., Knouri H., Gill J., Mikula A.,
 RA Bishal W.;
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
 RT laboratory strains";
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 CC -1: SUBCELLULAR LOCATION: Integral membrane protein (Potential).
 CC -1: SIMILARITY: BELONGS TO THE MMBP FAMILY.
 CC ;
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AL021928; CAA17334.1; -
 DR EMBL: AE006930; AAK44433.1; -
 DR TIGR: MT0212; -
 DR TuberculList; RV0202c; -
 DR InterPro; IPR000731; HMGCR/patch_5TM.
 DR PROSITE; PS50156; SSD; 1.
 KM Hypothetical protein; Transmembrane; Complete proteome.
 FT TRANSMEM 13 33 POTENTIAL.
 FT TRANSMEM 188 208 POTENTIAL.
 FT TRANSMEM 214 234 POTENTIAL.
 FT TRANSMEM 235 255 POTENTIAL.
 FT TRANSMEM 279 299 POTENTIAL.
 FT TRANSMEM 311 331 POTENTIAL.
 FT TRANSMEM 373 393 POTENTIAL.
 FT TRANSMEM 409 429 POTENTIAL.
 FT TRANSMEM 527 547 POTENTIAL.
 FT TRANSMEM 557 577 POTENTIAL.
 FT TRANSMEM 595 615 POTENTIAL.
 FT TRANSMEM 646 666 POTENTIAL.
 FT TRANSMEM 668 688 POTENTIAL.
 SQ SEQUENCE 966 AA; 103502 MW; 3E8BF0327CBEA2DA CRC64;

Alignment Scores:

Pred. No.: 0.0445 Length: 966
 Score: 121.50 Matches: 110
 Percent Similarity: 33.7% Conservative: 92
 Best Local Similarity: 18.06% Mismatch: 210
 Query Match: 3.67% Indels: 197
 DB: 1 Gaps: 25

US-09-868-987-1 (1-1864) x MMBP_MYCTU (1-966)

QY 281 CAAGCATTTCTCAGATCTCTGGCTTG----- 310
 Db 163 GlnGlyValLeuSerAlaAlaAlaAlaAsnThrLysHisAspIleAlaAsnAlaGlu 182
 QY 311 -----GCATGCTTATGTTTGTGATGAGCGTATATATATATGATTTGGAGGC 355
 Db 183 ArgTyrAsnLeuProIleIleLeuMetValLeuValAlaPhe-----Gly 198
 QY 356 GTCATCGCTTCGGAGCTGTTCTTCGATCTT-----TTGCTTATCTGGGCA 403
 Db 199 SerLeuAlaAlaAlaAlaAlaIleProLeuAlaLeuAlaValCysThrValIleThrMet 218
 QY 404 GCTCTACAGAT---TTGATGCGCCACTACCTTCGAGAGCTCGGTGGATTTGTTCTT 460
 Db 219 GlyLeuValAlaPheValLeuSerMetHisThrThrMetSerValPheValThrSerThrVal 238
 QY 461 GCTATG-----GGATGCGCGTAGATCAAAAGTTCTGTATTCGAAAGATCCGA 511
 Db 239 SerMetHeGlyIleAlaAlaValAlaAspTyrSerLeuPheIleLeuMetArgTyrArg 258
 QY 512 GAGCAATTTTATTTGCTCAAGCTTTTAAAAAATCTGTAGAAAAGATATACCAAGCT 571
 Db 259 GluGluLeuArgCysGlyArgArgProPheAspAlaValAlaPheAlaMetAlaThrSer 278
 QY 572 TTGGAGCCATTTTGTATTTCTTAACTTGACTACAGATATGGCCCTCAGACCTCTTCTTC 631
 Db 279 GlyLeuAlaValAlaLeuSerGlyMetThrValIleAlaSerLeuThrGlyIleTyrLeu 298
 QY 632 CTGATATCAGAGGCTTAAAGGTTTGTGACATGATTTTAAAGATTTTCTCTCA 691
 Db 299 IleAsnThrProAlaLeuArgSerMetAlaThrGlyAlaIleLeuAlaValAlaAla 318
 QY 692 ATGTTAGCGCTCTTTCATGACT----- 715
 Db 319 MetLeuThrSerAlaThrLeuThrProAlaValLeuAlaThrPheAlaArgAlaAlaAla 338
 QY 716 ---AAATTTTCTCATGCTGCTGATGATTAAGACCAATATACAGTTGATATGATG 772
 Db 339 LysArgSerAlaLeuValHisThrSerArgArgProAlaSerThrGlnSerThrPheTyr 358
 QY 773 AATAAGTTCGTGGGATTAAGCATGATTTCTTGAGAGATGCAAAAACCTTGG----- 826
 Db 359 SerArgTyrValGlyTyrPval-----MetArgArgProThrIleThr 372
 QY 827 GCTGTTTCGAAAGCTTTTCTTTAGTGGCTGCT-----CTC 868
 Db 373 AlaLeuAlaAlaSerThrValLeuLeuValMetAlaAlaProAlaThrLeuMetValLeu 392
 QY 869 GGGTTTGAGACCTGGAATTCCTGTTGGGAATGATTTTAAAGAGGATAGCTTTACC 928
 Db 393 Gly-----AsnSerLeuLeu---ArgGlnPheSerSerHisGluIle--- 406
 QY 929 TTATATCAAAAGAGCATGCGATCAGATGTTGCTCAATGCGTGAC----- 976
 Db 407 -----ArgThrGlyAlaAlaAlaAlaAlaGlnAlaLeuGlyProGlyAlaLeu 422
 QY 977 -----AAAGTTGCTAAACTACAGAAAGCTGCTTTCTTCTAGACCTTCGCT 1027
 Db 423 GlyProValGlnValLeuValArgPheAspAlaGlyGlyAlaSerAlaProGlu---His 441
 QY 1028 ATTCAAAATTTGGATCTTCAAGAAAGATCAAAATCTATTTTGTGATTAAGCTTTAAGC 1087
 Db 442 SerGlnThrIleAlaAla----- 447
 QY 1088 TATATCAAGAGATACAGAGCTCTCTCTTAAATTAACATGATGAGCTGCGTTATGCT 1147
 Db 448 -----IleArgHisArgIleAlaGlnAlaPro----- 456
 QY 1148 GGGATTTGTTCTCAAGAAAGAGCTTATTTCTTACGGAACCTTAA----- 1195
 Db 457 AsnValValSerValAlaProProAlaArgPheAlaAspAspAsnGlySerAlaLeuLeuSer 476

```
Qy 1196 -----CGAAGCGCAAAATTTTGGTCAAG 1219
Db 477 AlaValLeuSerValAspProGluAspLeuGlyAlaArgAspThrIleThrTrpMetArg 496
Qy 1220 GTAAGC-----AGCAAACTA 1234
Db 497 ThrGlnLeuProArgValAlaGlyAlaAlaGlnValAspValGlyGlyProThrAlaLeu 516
Qy 1235 TCGAAGAAATG-----CGTTATCAGCGGACC-----ATCGGGCTTTTGA 1276
Db 517 IleLysAspPheAspArgValSerAlaThrGlnProLeuValLeuValPheValAla 536
Qy 1277 GCTTTGGCAATCATCTGCTATGTGAGTTGCGC-----TTTCAA 1312
Db 537 ValIleAlaPheLeuMetLeuLeuSerIleArgSerValPheLeuAlaPheLysGly 556
Qy 1313 -----TTTCAA 1318
Db 557 ValLeuMetThrLeuLeuSerValAlaAlaAlaTyrGlySerLeuValMetValPheGln 576
Qy 1319 TGGCAATATGCTTTCAGTCCGCGTATGCGCTTAAATTCATGACCTTTTGGCTACCTGTGCA 1378
Db 577 TrpGlyTyrAlaArgGlyLeuGlyPheProAlaLeuHisSerIleAspSerThrValPro 596
Qy 1379 GCTTGTATTAGCAATCTCTTTTGAAGAAATCAAATAGATTGCAAGCCATTGGT 1438
Db 597 ProLeuValLeuAlaMetThrPhe-----604
Qy 1439 GCTTTAATGACTGTATGCGGTATTCATTAACAATACATTTTGATCATTTTGATCGTATT 1498
Db 605 -----GlyLeuSerMetAspTyrGluIlePheLeuLeuThrArgIle 618
Qy 1499 CGTGAAGATCGCAAGCAACCTGTTTACCCCTATGCATGTTTATGTTTAAATGATGCCCTT 1558
Db 619 ArgGlu-----ArgPheLeuGlnThrGlyGlnThr 628
Qy 1559 CAAAGAGCTTCAGCGGCGGATGATGACAGCAGCTACAACCTCA-----TCAGTTTGT 1612
Db 629 ArgAspAlaValAlaTyrGlyValArgThrSerAlaArgThrIleThrSerAlaAlaLeu 648
Qy 1613 TTAATGCTTTTGTATAGCGGCTCCTGCTTAAATTTTGCATTATATGACCATA 1672
Db 649 IleMetIleAlaValPheCysGlyPheAlaPheAlaGlyMetProLeuValAlaGluIle 668
Qy 1673 GGGATCTTCTCA-----GGAATCTTATCGTCTCTTTATATGCA 1711
Db 669 GlyValAlaCysAlaValAlaIleAlaValAspAlaThrValValArgLeuValLeuVal 688
Qy 1712 CCACCTCTGTTGTTTATGTCGCT 1738
Db 689 ProAlaLeuMetAlaMetPheAspArg 697
RESULT 45
ACRD_ECOLI STANDARD; PRT; 1037 AA.
ID P24177; P77178; Q46715; P76971;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable aminoglycoside efflux pump (Acriflavine resistance protein
DE D).
GN ACRD OR B2470.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia
OX NCBI_Taxid=562;
RN
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RA Nilles M.L., Bertrand K.P.;
RL Submitted (JUL-1994) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
```

```
RA Ma D., Cook D.N., Alberti M., Nikaïdo H., Hearst J.E.;
RL Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RN Science 277:1453-1474(1997).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97349980; PubMed=9205837;
RA Yamamoto Y., Alba H., Baba T., Hayaashi K., Inada T., Isono K.,
RA Itoh T., Kimura S., Kitagawa M., Makino K., Miki T., Mitsuhashi N.,
RA Mizobuchi K., Mori H., Nakade S., Nakamura Y., Nashimoto H.,
RA Oshima T., Oyama S., Saito N., Sampei Y., Sato Y., Sivasubram S.,
RA Tagami H., Takahashi H., Takeda J., Takemoto K., Uehara K., Wada C.,
RA Yamagata S., Horiuchi T.;
RT "Construction of a contiguous 874-kb sequence of the Escherichia coli
RT - K12 genome corresponding to 50.0-88.8 min on the linkage map and
RT analysis of its sequence features.";
RN DNA Res. 4:91-113(1997).
RN [5]
RP SEQUENCE OF 998-1037 FROM N.A.
RC STRAIN=K12;
RX MEDLINE=92355499; PubMed=1644752;
RA Bouvier J., Richard C., Higgins W., Bogler O., Stragier S.;
RT "Cloning, characterization, and expression of the dapE gene of
RT Escherichia coli.";
RN J. Bacteriol. 174:5265-5271(1992).
RN [6]
RP FUNCTION.
RX MEDLINE=20158894; PubMed=10692383;
RA Rosenberg E.Y., Ma D., Nikaïdo H.;
RT "AcrD of Escherichia coli is an aminoglycoside efflux pump.";
RN J. Bacteriol. 182:1754-1756(2000).
CC -I- FUNCTION: PARTICIPATES IN THE EFFLUX OF AMINOGLYCOSIDES. CONFERS
CC RESISTANCE TO A VARIETY OF THESE SUBSTANCES.
CC -I- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC (By similarity).
CC -I- SIMILARITY: BELONGS TO THE ACRB/ACRD/ACRF FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/ebis-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
```

```

FT DOMAIN 360 365 CYTOPLASMIC (BY SIMILARITY).
FT TRANSMEM 366 385 3 (BY SIMILARITY).
FT DOMAIN 386 391 PERIPLASMIC (BY SIMILARITY).
FT TRANSMEM 392 413 4 (BY SIMILARITY).
FT DOMAIN 414 441 CYTOPLASMIC (BY SIMILARITY).
FT TRANSMEM 442 460 5 (BY SIMILARITY).
FT DOMAIN 461 473 PERIPLASMIC (BY SIMILARITY).
FT TRANSMEM 474 496 6 (BY SIMILARITY).
FT DOMAIN 497 537 CYTOPLASMIC (BY SIMILARITY).
FT TRANSMEM 538 556 7 (BY SIMILARITY).
FT DOMAIN 557 570 PERIPLASMIC (BY SIMILARITY).
FT TRANSMEM 571 590 8 (BY SIMILARITY).
FT DOMAIN 591 596 CYTOPLASMIC (BY SIMILARITY).
FT TRANSMEM 597 616 9 (BY SIMILARITY).
FT DOMAIN 617 622 PERIPLASMIC (BY SIMILARITY).
FT TRANSMEM 623 644 10 (BY SIMILARITY).
FT DOMAIN 645 671 CYTOPLASMIC (BY SIMILARITY).
FT TRANSMEM 672 690 11 (BY SIMILARITY).
FT DOMAIN 691 1003 PERIPLASMIC (BY SIMILARITY).
FT TRANSMEM 1004 1026 12 (BY SIMILARITY).
FT DOMAIN 1027 1037 CYTOPLASMIC (BY SIMILARITY).
FT TRANSMEM 1038 1037 A -> G (IN REF. 1).
FT CONFLICT 303 303 V -> E (IN REF. 1).
FT CONFLICT 372 372 A -> D (IN REF. 2).
FT CONFLICT 385 385 G -> P (IN REF. 1).
FT CONFLICT 461 461 S -> PD (IN REF. 1).
FT CONFLICT 665 665 R -> A (IN REF. 1).
FT CONFLICT 763 763 A -> G (IN REF. 1).
FT CONFLICT 775 775 R -> P (IN REF. 1).
FT CONFLICT 778 778
SQ SEQUENCE 1037 AA; 113047 MW; 961611E1D24FD4E5 CRC64;

```

```

Alignment Scores:
Pred. No.: 0.0783 Length: 1037
Score: 118.50 Matches: 146
Percent Similarity: 30.26% Conservative: 97
Best local Similarity: 18.18% Mismatches: 267
Query Match: 3.58% Indels: 293
DB: 1 Gaps: 33

```

US-09-868-987-1 (1-1864) x ACRD_ECOLI (1-1037)

```

QY 44 AATGGAATTTCTGCAACCGTGAATGGCTGATGAGTATGACGGTATATG 103
   |||||:::|||||:::|||||:::|||||:::
Db 282 AsnGlySerProAlaSerGlyLeuGlyValLeuAlaSerGlyAlaAsnGlyMetAla 301
   :::::|||||:::|||||:::|||||:::
QY 104 GTCAGCAGCCATTATTTAAACGTCCTGAAATCATGCGACAGTCTCGAGAAATT 163
   :::::|||||:::|||||:::|||||:::
Db 302 ThrAlaGluLeuValLeuAsn-----ArgLeuAspGluLeuAlaGlnTyrPhe 317
   :::::|||||:::|||||:::|||||:::
QY 164 ACCCAGCGTGAAGTGAACAACCTGCGCTCAGATTAAATCTGAGCGATCTTTGTT 223
   |||||:::|||||:::|||||:::|||||:::
Db 318 ProHseGlyLeuGlnTyrValAlaTyrThrSerPheValTyrAlaSerIle 337
   :::::|||||:::|||||:::|||||:::
QY 224 CCCAGTTCCTCAGTGAAGAGACGATCTCTGATCTTGGGAAAAAACAATGACAA 283
   :::::|||||:::|||||:::|||||:::
Db 338 GluAspValVal-----LysThrLeuLeuGln 346
   :::::|||||:::|||||:::|||||:::
QY 284 GGCATTATCTCAGCATGCTGTGGCTTGCAATGCTTATTTGTTGATGAGCGATATATTAT 343
   |||||:::|||||:::|||||:::|||||:::
Db 347 AlaIle-----AlaLeuValPheLeuValMetTyrLeuPheLeuGlnAsnPheArgAla 364
   :::::|||||:::|||||:::|||||:::
QY 344 AGATTGGAGGCGTCAATGCTTCGCGAGCGTCTTCTGAAATCTTTGCTTATCTGGCA 403
   |||||:::|||||:::|||||:::|||||:::
Db 365 ThrLeuIleProThrIleAlaValProValLeuMetGlyThrPheSerValLeuTyr 384
   :::::|||||:::|||||:::|||||:::
QY 404 GCTTACAGTATTTGATGATGCCCACTGTCAGACGTCGCGGATTTGTTGCTTGGCT 463
   |||||:::|||||:::|||||:::|||||:::
Db 385 AlaHseGlyTyrSerValAsnThrLeuThrMetPheAlaMet-----ValLeuAla 401
   :::::|||||:::|||||:::|||||:::
QY 464 ATGGGATGGCGGTAGATCAATATGTTCTTATTCGAAAGATCCGAGAGAAATTTTGA 523
   |||||:::|||||:::|||||:::|||||:::
Db 402 IleGlyLeuLeuValAlaSerPheAlaIleValValGluAsnVal---GluArgIleMet 420

```

```

QY 524 TTGTCGAAGTCTT-----AAAAATCGTGAAGAAAGATATACCAAGCTTTTGA 577
   ::|||:::|||||:::|||||:::
Db 421 SerGluGlnIleLeuThrProArgGluAlaThrArgLysSerMetGlyGlnIleGlnGly 440
   |||||:::|||||:::|||||:::
QY 578 GCCATT-----TTGATTCCTAATCTGACATGATGATGTCCTCAGACACTCTTTCTTC 631
   ::|||:::|||||:::|||||:::
Db 441 AlaLeuValGlyIleAlaMetValLeuSerAlaValPheValProMetAlaPhePheGly 460
   ::|||:::|||||:::|||||:::
QY 632 CTGGAATACAGAGCGCTTAT--AAAGGTTTGCTTGACATGATTTAGGAATTTCTCT 688
   |||||:::|||||:::|||||:::
Db 461 GlyThrThrGlyAlaIleTyrArgGlnPheSerIleThrIleValAlaAlaMetValLeu 480
   ::|||:::|||||:::|||||:::
QY 689 TCATGTTTACGGCTCTTTTCATGACTGAATTTTC----- 724
   |||||:::|||||:::|||||:::
Db 481 SerValLeuValAlaMetIleLeuThrProAlaLeuCysAlaThrLeuLeuLysProLeu 500
   ::|||:::|||||:::|||||:::
QY 725 -----TTCATGCTGTGATGAATAGACCAACAT 754
   |||||:::|||||:::|||||:::
Db 501 LysLysGlyGluHisIleGlyGlnLysGlyPhePheAlaThrPheAsnGlnMetPheAsn 520
   ::|||:::|||||:::|||||:::
QY 755 ACAAGTTCATATGATGAATTAAGTTCGTGGG--ATTAAGCATGATTTCTGAGAGCA 811
   |||||:::|||||:::|||||:::
Db 521 ArgAsnAlaGluArgTyrGlyLysGlyValAlaLysIleLeuHisArgSerLeuArg--- 539
   ::|||:::|||||:::|||||:::
QY 812 TGCAAAAACTTTGGGCTGTT-----TCTGGAAGTCTTTCTT--- 850
   |||||:::|||||:::|||||:::
Db 540 -----ThrIleValIleTyrValLeuLeuLeuGlyGlyMetValPheLeuPhe 555
   ::|||:::|||||:::|||||:::
QY 850 ----- 850
   |||||:::|||||:::|||||:::
Db 556 LeuArgLeuProThrSerPheLeuProLeuGluAspArgGlyMetPheThrThrSerVal 575
   ::|||:::|||||:::|||||:::
QY 850 ----- 850
   |||||:::|||||:::|||||:::
Db 576 GlnLeuProSerGlySerThrGlnGlnThrLeuLysValGluGlnIleGluLys 595
   ::|||:::|||||:::|||||:::
QY 851 -----TTAGTTCGCTGCTCGCGG 871
   |||||:::|||||:::|||||:::
Db 596 TyrTyrPheThrHisGluLysAspAsnIleMetSerValPheAlaThrValGlySerGly 615
   ::|||:::|||||:::|||||:::
QY 872 TTTGAGCC-----TGAAATTC 889
   |||||:::|||||:::|||||:::
Db 616 ProGlyGlyAsnGlyGlnAsnValAlaArgMetPheIleArgLeuLysAspTyrPheSerGlu 635
   ::|||:::|||||:::|||||:::
QY 890 GTTTGGAAATGATTTAAAGAGGCTATGCTTTACCTT----- 931
   |||||:::|||||:::|||||:::
Db 636 -----ArgAspSerLysThrGlyThrSerPheAlaIleIleGluArgAlaThrLys 652
   ::|||:::|||||:::|||||:::
QY 931 ----- 931
   |||||:::|||||:::|||||:::
Db 653 AlaPheAsnGlnIleLysGluAlaArgValIleAlaSerSerProProAlaIleSerGly 672
   ::|||:::|||||:::|||||:::
QY 931 ----- 931
   |||||:::|||||:::|||||:::
Db 673 LeuGlySerSerAlaGlyPheAspMetGluLeuGlnAspHisAlaGlyAlaGlyHisAsp 692
   ::|||:::|||||:::|||||:::
QY 932 -----AATCCA----- 937
   |||||:::|||||:::|||||:::
Db 693 AlaLeuMetAlaAlaArgAsnGlnLeuLeuAlaLeuAlaIleGluAsnProGluLeuThr 712
   ::|||:::|||||:::|||||:::
QY 938 -----AAAAGATGATGATCAGCAGCGATGTCCTCAAAATGGGTGGCAAAAGTT---GTGCAT 988
   ::|||:::|||||:::|||||:::
Db 713 ArgValArgHisAsnGlyLeuAspAspSerProGlnLeuGlnIleAspIleAspGlnArg 732
   ::|||:::|||||:::|||||:::
QY 989 AAATCAGAGAACCTGCTCTTCTTCTAGACATTCCTG-----ATTCAAACA---TTT 1039
   |||||:::|||||:::|||||:::
Db 733 LysAlaGlnAlaLeuGlyValAlaAlaIleAspAspIleAsnAspThrLeuGlnThrAlaTyr 752
   ::|||:::|||||:::|||||:::
QY 1040 GGATCTTCA-----GAAAAGTCAAAATCTATTTTACTGAT 1075
   |||||:::|||||:::|||||:::
Db 753 GlySerSerTyrValAsnAspPheMetAspArgGlyArgValLysValTyrValGln 772
   ::|||:::|||||:::|||||:::
QY 1076 AAAGCTTAAGCTATCTAGACAGATACGAGCTCTCTCTAAATTAACGATCATGAGC 1135

```


Db	773	AlaAlaAlaProTyr			----	ArgMetLeuProAspAspIleAsnLeu	786
Qy	1136	TGCGGTTATTGTGGGATTGTGTGAGAAACAGGCCCTAGATTCTCTACGGAAACTCTAAA			----	ArgMetLeuProAspAspIleAsnLeu	1195
Db	787	Trp	-----	TyrValArgAsnIysAspGlyGlyMetValProPheSerAla	801		
Qy	1196	CGAAACGCAAAATTTTGGTCAAAAGGTAAAGCAGCAACATTCGAAAGAAAATCGCTTATCAG			----	ArgMetLeuProAspAspIleAsnLeu	1255
Db	802	PheAlaThrSerArgTrpGluThrGlySerProArgLeuGluArgTyrAsnGlyTyrSer			----	ArgMetLeuProAspAspIleAsnLeu	821
Qy	1256	GCAGCC	-----	-----	-----	ArgMetLeuProAspAspIleAsnLeu	1261
Db	822	AlaValGluIleValGlyGluAlaAlaProGlyValSerThrGlyThrAlaMetAspIle	841				
Qy	1262	-----	ATCGGGCTT	-----	-----	ArgMetLeuProAspAspIleAsnLeu	1270
Db	842	MetGluSerLeuValIysGlnLeuProAsnGlyPheGlyLeuGluTrpThrAlaMetSer	861				
Qy	1271	-----	-----	TTAGGAGCTTTGGCA	1285		
Db	862	TyrGlnGluArgLeuSerGlyAlaGlnAlaProAlaLeuTyrAlaIleSerLeuLeuVal	881				
Qy	1286	ATCATCTGCTCTATGTGAGTTGGCTTTGAAATGCGCAATATGCTTTCAGTCCGTA	--- 1342				
Db	882	ValPheLeuCysLeuAlaLeuTyrGluSerTrpSerValProPheSerValMetLeu	901				
Qy	1343	-----	TGCGCTTTAATTCATGACCTTTTGGCTACTGTGCGACTCTTGTATTATAGCA	1393			
Db	902	ValValProLeuGlyValIleGlyAlaLeuLeuAlaThr	-----	-----	-----	ArgMetLeuProAspAspIleAsnLeu	914
Qy	1394	CATTCTTTTGAAGAAATTCAAATAGATTGTCAGCCATGTGTGCTTTAAGACTGTA	1453				
Db	915	-----	TrpMetArgGlyLeuGluAsnAspValTyrPheGlnValGlyLeuLeuThrVal	932			
Qy	1454	TTGGGTATTCAATTAACATACTTTGATCATTTTGTATCGTATTCGTGAGATCGCAA	1513				
Db	933	IleGlyLeuSerAlaIysAsnAlaIleuIleValGluPheAlaAsnGluMetAsnGln	952				
Qy	1514	GCG	-----	AACCTGTTT	-----	ArgMetLeuProAspAspIleAsnLeu	1564
Db	953	LysGlyHisAspLeuPheGluAlaThrLeuHis	-----	-----	-----	ArgMetLeuProAspAspIleAsnLeu	967
Qy	1565	ACGTTCAGCCGCAAGTAATGACAACA	--- GCTACA	CTCTATCAGTTTGTGTTAATGCTT	1621		
Db	968	ArgLeuArgProIleLeuMetThrSerLeuAlaPheIlePheGlyValLeuProMetAla	987				
Qy	1622	TTGTTTATAGCGGCTCTGCTCTTAATTTTGCATTATTATGACCATAGGATTCCT	1681				
Db	988	ThrSerThrGlyAlaGlySerGlyGlnHisAla	-----	ValGlyThrGlyValMet	1005		
Qy	1682	CTAGGAATTTATCGTCTTTTATATGCA	-----	CCACCTCTGTTGTTGTTTATG	1732		
Db	1006	GlyGlyMetIleSerAlaThrIleLeuAlaIleTyrPheValProLeuPheValLeu	1025				
Qy	1733	GTCCGTAAA	1741				
Db	1026	ValArgArg	1028				

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: December 17, 2002, 12:38:47 ; Search time 19.5 Seconds
(without alignments)
3185.498 Million cell updates/sec

Title: US-09-868-987-1

Perfect score: 3312

Sequence: 1 atggaactccgattatgtc.....atgacaaattcagataatgc 1864

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 105981 seqs, 16662342 residues

Total number of hits satisfying chosen parameters: 211962

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool/US09868987/runat 16122002 083009 23678/app query.fasta_1.2055
-DB=Published Applications AA -QFMT=faetan -SUFFIX=rapb -MINMATCH=0.1
-LOPCL=0 -LOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=45 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US09868987 @C@N 1 1 5 -tunat 16122002 083009 23678
-NCPU=6 -ICPU=3 -NO XLPXY -NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications AA:*

1: /cgn2_6/ptodata/2/pubpaa/US08 NEW PUB.pcp.*
2: /cgn2_6/ptodata/2/pubpaa/PCT NEW PUB.pcp.*
3: /cgn2_6/ptodata/2/pubpaa/US06 NEW PUB.pcp.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pcp.*
5: /cgn2_6/ptodata/2/pubpaa/US07 NEW PUB.pcp.*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pcp.*
7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pcp.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pcp.*
9: /cgn2_6/ptodata/2/pubpaa/US09_NEW PUB.pcp.*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pcp.*
11: /cgn2_6/ptodata/2/pubpaa/US10 NEW PUB.pcp.*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pcp.*
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW PUB.pcp.*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pcp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	2778	83.9	552	10	US-09-886-468-14
2	678.5	20.5	737	10	US-09-899-482-2
3	431	13.0	525	10	US-09-881-752A-96
4	364	11.0	323	10	US-09-899-482-4

5	359.5	10.9	615	10	US-09-899-482-3	Sequence 3, Appli
6	277.5	8.4	637	10	US-09-852-053-3	Sequence 3, Appli
7	245.5	7.4	279	10	US-09-861-451A-28	Sequence 28, Appl
8	228	6.9	403	10	US-09-852-053-4	Sequence 4, Appli
9	147.5	4.5	1055	10	US-09-815-242-12189	Sequence 12189, A
10	144	4.3	1034	10	US-09-815-242-10331	Sequence 10331, A
11	136.5	4.1	1032	10	US-09-815-242-11145	Sequence 11145, A
12	129	3.9	1040	10	US-09-912-020-282	Sequence 282, App
13	129	3.9	1040	10	US-09-815-242-10203	Sequence 10203, A
14	128	3.9	1028	10	US-09-815-242-11516	Sequence 11516, A
15	123.5	3.7	1025	10	US-09-912-020-283	Sequence 283, App
16	122.5	3.7	1062	10	US-09-815-242-5111	Sequence 5111, Ap
17	122	3.7	761	9	US-09-895-913A-10	Sequence 10, Appl
18	122	3.7	1043	10	US-09-815-242-5194	Sequence 5194, Ap
19	118	3.6	801	10	US-09-815-242-5330	Sequence 5330, Ap
20	117.5	3.5	1049	10	US-09-815-242-10072	Sequence 10072, A
21	116.5	3.5	1018	10	US-09-815-242-5186	Sequence 5186, Ap
22	111.5	3.4	467	10	US-09-767-041-23	Sequence 23, Appl
23	111	3.5	597	12	US-10-139-876-20	Sequence 20, Appl
24	109	3.3	396	10	US-09-815-242-10968	Sequence 10968, A
25	108	3.3	651	10	US-09-815-242-5331	Sequence 5331, Ap
26	108	3.3	651	10	US-09-815-242-12682	Sequence 12682, A
27	108	3.3	1049	10	US-09-815-242-13719	Sequence 13719, A
28	106	3.2	488	12	US-10-094-059-4	Sequence 4, Appli
29	102.5	3.1	487	9	US-10-095-139-14	Sequence 14, Appl
30	102.5	3.1	487	10	US-09-795-693-27	Sequence 27, Appl
31	101	3.0	381	10	US-09-827-937A-18	Sequence 18, Appl
32	101	3.0	381	10	US-09-842-316-6	Sequence 6, Appli
33	101	3.0	381	10	US-09-731-030A-17	Sequence 17, Appl
34	99.5	3.0	881	10	US-09-982-736-2	Sequence 2, Appli
35	99.5	3.0	1053	10	US-09-815-242-5136	Sequence 5136, Ap
36	98.5	3.0	396	10	US-09-741-669-395	Sequence 395, App
37	97.5	2.9	382	10	US-09-969-711-2	Sequence 13, Appli
38	96.5	2.9	341	9	US-09-971-228-13	Sequence 15, Appl
39	96.5	2.9	346	9	US-09-971-228-15	Sequence 8, Appli
40	96.5	2.9	362	10	US-09-993-844-8	Sequence 19, Appl
41	96.5	2.9	381	12	US-10-037-616-19	Sequence 5, Appli
42	96.5	2.9	382	9	US-09-971-228-5	Sequence 7, Appli
43	96	2.9	486	10	US-09-860-232A-7	Sequence 4, Appli
44	96	2.9	619	10	US-09-741-153-4	Sequence 14, Appl
45	94.5	2.9	314	9	US-09-971-228-14	

ALIGNMENTS

RESULT 1
US-09-886-468-14
; Sequence 14, Application US/09886468
; Patent No. US20020037293A1
; GENERAL INFORMATION:
; APPLICANT: Aventis Pasteur Limited
; TITLE OF INVENTION: Chlamydia antigens and corresponding DNA fragments and uses thereof
; FILE REFERENCE: 77813-5
; CURRENT APPLICATION NUMBER: US/09/886,468
; CURRENT FILING DATE: 1999-12-23
; PRIOR APPLICATION NUMBER: 60/113,280
; PRIOR FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: 60/113,281
; PRIOR FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: 60/113,282
; PRIOR FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: 60/113,283
; PRIOR FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: 60/113,284
; PRIOR FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: 60/113,285
; PRIOR FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: 60/113,385
; PRIOR FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: 60/114,050
; PRIOR FILING DATE: 1998-12-28
; PRIOR APPLICATION NUMBER: 60/114,056
; PRIOR FILING DATE: 1998-12-28

; PRIOR APPLICATION NUMBER: 60/114,057
 ; PRIOR FILING DATE: 1998-12-28
 ; PRIOR APPLICATION NUMBER: 60/114,058
 ; PRIOR FILING DATE: 1998-12-28
 ; PRIOR APPLICATION NUMBER: 60/114,059
 ; PRIOR FILING DATE: 1998-12-28
 ; PRIOR APPLICATION NUMBER: 60/114,061
 ; PRIOR FILING DATE: 1998-12-28
 ; NUMBER OF SEQ ID NOS: 26
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 14
 ; LENGTH: 552
 ; TYPE: PRF
 ; ORGANISM: Chlamydia pneumoniae
 ; US-09-868-468-14

Alignment Scores:

Prod. No.:	1,57e-274	Length:	552
Score:	2773.00	Matches:	552
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	83.38%	Indels:	0
DB:	10	Gaps:	0

US-09-868-987-1 (1-1864) x US-09-868-468-14 (1-552)

QY 101 ATGTCGACGACCCCTATTTTAAAGTCCCATGAAAAATCATGCCAGTGTCTCAGGAAA 160
 Db 1 MetValSerSerProIleLeuAsnValProLeuYsAsnHisAlaSerValSerGlyLys 20
 QY 161 TTTTACCACCGTGAAGTACGAAACCTGGCCCTCAGATTAAATCTTGAGAGCATCTTTT 220
 Db 21 PheThrIstArgGluValSerLysLeuAlaSerAspLeuLysSerGlyAlaMetSerPhe 40
 QY 221 GTTCCCGAGGTTCTCAGTGAAGAGACGATCTCTCTGATCTTGAGAAAAACAATGTACA 280
 Db 41 ValProIleValLeuSerGluGluThrIleSerSerAspLeuGlyLysGlnCysThr 60
 QY 281 CAAGGCAATTAATCTCAGCATGCTGTGGCTTGCAATGCTTATTTGTTGATGAGGTAAT 340
 Db 61 GlnGlyIleIleSerAlaCysGlyLeuAlaMetLeuIleValLeuMetSerValTyr 80
 QY 341 TATATGATTTGGAGCGCTTCATGCTTGGAGACCTGTCTCTCGAATCTTTTGCTATCTGG 400
 Db 81 TyrArgPheGlyGlyValIleAlaSerGlyAlaValLeuLeuAsnLeuLeuIleTyr 100
 QY 401 GCGAGCTTACAGTATTTGGATGGCGCACTGCTGTGAGGACTCGGTGGAGTTGTTCTT 460
 Db 101 AlaIleLeuGlnIleValLeuAlaProLeuThrLeuSerGlyLeuAlaGlyIleValLeu 120
 QY 461 GCTATGGGATGGCGGTAGATGCAATGTTCTTGTATTCGAAGAATCCGAGAGAAATTT 520
 Db 121 AlaeMetGlyMetAlaValAspAlaAsnValLeuValPheGluArgIleArgGluGluPhe 140
 QY 521 TTTATGTTCTCAAGTCTTAAAAAATCTGTAGAAAAAGATATACCAAGGCTTTTGGAGCC 580
 Db 141 LeuLeuSerGlnSerIleuLysValSerValGluLysGlyTyrThrLysAlaPheGlyAla 160
 QY 581 ATTTTGAATTTAGTACTGACAGTATTTGGCTCAGACCTCTTTCTTCTTGATGATACA 640
 Db 161 IlePheAspSerAsnLeuThrThrValLeuAlaSerAlaLeuLeuPhePheLeuAspThr 180
 QY 641 GGGCCTATTAAAGGGTTGCTTTGACATTTGATTTTGAAGATTTTCTCTCAATGTTTACG 700
 Db 181 GlyProIleLysGlyPheAlaLeuThrLeuIleLeuGlyIlePheSerSerMetPheThr 200
 QY 701 GCTCTTTTTCATGACTAAATTTTCTTCATGCTGTGATGAATTAAGACCAACATACACAG 760
 Db 201 AlaLeuPheMetThrLysPhePhePheMetLeuThrPheAsnLysThrGlnHisThrGln 220
 QY 761 TTGATATGATGATTAAGTTCTGTGGGGTAAAGCATGTTTCTTGAGAGATGAGAAAAA 820
 Db 221 LeuHisMetMetAsnLysPheValGlyIleLysHisAspPheLeuArgGlyCysLeuLys 240

QY 821 CTTTGGGCTGTCTTCGGAAGTGTCTTTCTTTAGTTGGCTTGTCTCGGGTTTGGAGCC 880
 Db 241 LeuThrAlaValSerIleLysSerValPheLeuLeuGlyCysValAlaLeuGlyPheGlyAla 260
 QY 881 TGGATTTCCGTTTGGGAATGATTTTAAAGAGAGGATATGACCTTTACCTTTAATCCAAA 940
 Db 261 TrpAsnSerValIleuGlyMetAspPheLysGlyGlyTyrAlaPheThrPheAsnProLys 280
 QY 941 GAGCATGGCATCAGCGATGTGCTCAATGCTGAGCAAAAGTTGTGATTAATACAGAA 1000
 Db 281 GlnHisGlyIleSerAspValAlaGlnMetArgGlyLysValValHisLysLeuGlnGlu 300
 QY 1001 GCTGTCTCTTTCTTCAGACATCTCCGATTCACAACTTTGATCTTCAGAAAAAGATCAA 1060
 Db 301 AlaGlyLeuSerSerThrArgAspPheArgIleGlnThrPheGlySerSerGlyLysIleLys 320
 QY 1061 ATCTATTTTATGATTAAGCTTTTAACTTACTTACGATACGATGACGCTCTCTCAAAA 1120
 Db 321 IleTyrPheSerAspLysAlaLeuSerTyrThrLysGlnIleArgAlaSerLeuLeuLys 340
 QY 1121 TTAACGATCATGAGCTGGCGTTATTTGTGGATTTGTGCAGAAACAGGCTAGATTCTC 1180
 Db 341 LeuThrIleMetSerThrArgTyrCysGlyIleValValArgAsnArgProArgPheLeu 360
 QY 1181 TACGAAATCTTAAAGAAACGCAAAATTTTGTGCAAGGTAAAGCAACGCAATCTGAAAG 1240
 Db 361 TyrGlyAsnSerLysArgAsnAlaLysPheTyrSerLysValSerSerLysLeuSerLys 380
 QY 1241 AAATGCGTTATCAGCGCATCGGCGCTTTTGAAGCTTTGAGCAATCATCTGCTCTAT 1300
 Db 381 LysMetArgTyrGlnAlaThrIleGlyLeuLeuGlyAlaLeuAlaIleLeuLeuTyr 400
 QY 1301 GTGAGTTTGGCTTGAATGCAATATGCTTTCAGTCCGATACGCTTTAATTCATGAC 1360
 Db 401 ValSerLeuArgPheGluThrGlnTyrAlaPheSerAlaValCysAlaLeuIleHisAsp 420
 QY 1361 CTTTGGCTTACCTGTGACGCTTGTATTATGACATTTCTTTTGAAGAAAAATTCAAAT 1420
 Db 421 LeuLeuAlaThrCysAlaValLeuPheIleAlaHisPheLeuLysLysIleGlnIle 440
 QY 1421 GATTTGCAAGCCATGTTGTGTTAATGACTGATTTGGGGATTCATTAACATATCTTGG 1480
 Db 441 AspLeuGlnAlaIleGlyAlaLeuMetThrValLeuGlyTyrSerLeuAsnThrLeu 460
 QY 1481 ATCATTTTGTATGATTTGATGATGATCGCAAGATCGCAACGATGTTTACCCCTATGATGT 1540
 Db 461 IleIlePheAspArgIleArgIleAspArgGlnAlaAsnLeuPheThrPromethisVal 480
 QY 1541 TTTGTTAATGATGCTCTTCAAAAAGCTTCAAGCGGCAAGGTAATGACACAGCTACACT 1600
 Db 481 LeuValAsnAspAlaLeuGlnLysThrPheSerArgThrValMetThrThrAlaThrThr 500
 QY 1601 CTATCAGTTTGTATAGCTTTGTTTATAGGCGGCTCTGCTTATTTTATTTTGCATTT 1660
 Db 501 LeuSerValLeuLeuMetLeuLeuPheIleGlyGlySerSerValPheAsnPheAlaPhe 520
 QY 1661 ATTATGACATAGGATTTCTTACGAACTTTACGCTCTTATATTTGACACCACTCTG 1720
 Db 521 IleMetThrIleGlyIleLeuLeuGlyThrLeuSerSerLeuThrIleAlaProLeu 540
 QY 1721 TTGTTGTTTATGCTCGGTAAAGAAATCGCTCAAAA 1756
 Db 541 LeuLeuPheMetValArgLysGluAsnArgSerLys 552

RESULT 2

US-09-899-482-2
 ; Sequence 2, Application US/09899482
 ; Patent No. US2002006641A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Quax, Wilhelmus J.
 ; TITLE OF INVENTION: Increasing Production of Proteins in Microorganisms
 ; FILE REFERENCE: GC385-PC7

;; CURRENT APPLICATION NUMBER: US/09/899,482
;; CURRENT FILING DATE: 2001-07-05
;; PRIOR APPLICATION NUMBER: EP 97305286.3
;; PRIOR FILING DATE: 1997-07-16
;; PRIOR APPLICATION NUMBER: EP 97305344.0
;; PRIOR FILING DATE: 1997-07-17
;; NUMBER OF SEQ ID NOS: 4
;; SOFTWARE: FASTSEQ for Windows Version 3.0
;; SEQ ID NO 2
;; LENGTH: 737
;; TYPE: PRT
;; ORGANISM: Bacillus Subtilis
US-09-899-482-2

Alignment Scores:
Pred. No.: 8,92e-61 Length: 737
Score: 678.50 Matches: 169
Percent Similarity: 50.43% Conservative: 125
Best Local Similarity: 28.99% Mismatches: 192
Query Match: 20.49% Indels: 97
DB: 10 Gaps: 15

US-09-868-987-1 (1-1864) x US-09-899-482-2 (1-737)

Qy 104 GTCAGCAGCCCTATTAAAGCTCCATTGAAA---AATCATGCCAGTGTCTCAGGGAAA 160
Db 197 ValSerAlaProAsnValSerGlnGluLeuAenThrThrAspValLysLeuGluGlyHis 216
Qy 161 TTTACCCACCGTGAAGTGCAGAACTCCCTCAGATTTAAATCTGGAGCGATGCTTTT 220
Db 217 PhetrAlaGlnGluAlaLysAspLeuAlaSerlleLeuAenAlaGlyAlaLeu----- 234
Qy 221 GTTCCGAGGTTCTCAGTGAAGAG-----ACGATCTCTCTCTGATCTTTGGGAAAAA 271
Db 235 ---ProValLysLeuThrGluLysTyrSerThrSerValGlyAlaGlnPheGlyGlnGln 253
Qy 272 CAATGTACACAGGATATCTCAGCATGCTGTGGCTTGGCATGCTATTGTTTGTATG 331
Db 254 AlaLeuHisAspThrValPheAlaGlylleValGlylleAlaIlePheLeuPheMet 273
Qy 332 AGCGTATATTATAGATTGGAGGCGTCATCGCTTCGGGAGCTGTTCTCTGAATCTTTG 391
Db 274 LeupheTyrTyrArgLeuProGlyLeuIle-----AlaValilleThrLeuSerVal 290
Qy 392 CTTATCTGGCAGCTCTACAG-----TATTTGGATGGCCACTCACCTTGTTCAGGA 442
Db 291 TyrlleTyrilleThrLeuGlnIlePheAspTrpMetAsnAlaValLeuThrLeuProGly 310
Qy 443 CTCGCTGGATTTCTGCTATGGGATGGCCGATAGTGAATGTTCTGTATTCGAA 502
Db 311 lleAlaAlaLeuIleLeuGlyValGlyMetAlaValAspAlaAsnIlelleThrTyrGlu 330
Qy 503 AGAATCCGAGAGGATTTTATTGTCCTCAAGTCTTAAATAATCTGAGAAAAAGGATAT 562
Db 331 ArgilleLysGluLeuLeuLysLeuGlyLysSerValArgSerAlaPheArgSerGlyAsn 350
Qy 563 ACCAAGGCTTTGGAGCCATTTTTCATCTCACTTACATGATTCAGTATGGCTCTCAGACT 622
Db 351 ArgArgSerPheAlaThrIlePheAspAlaAsnIleThrThrIlelleAlaAlaValVal 370
Qy 623 CTTTTCCTCCTAGATACAGGGCTATTAAAGGTTTGTCTTGGACATTCATTTAGGAAT 682
Db 371 LeuPheillePheGlyThrSerSerValLysGlyPheAlaThrMetLeuIleLeuSerille 390
Qy 683 TTCCTCTCAATGTTTACGGCTTTTTCATGACTAAATTTTCTTCATGCTG----- 733
Db 391 LeuThrSerPheilleThrAlaValPheLeuSerArgPheLeuLeuAlaLeuValGlu 410
Qy 734 -----TGCATGAATAAG----- 745
Db 411 SerArgTTPLeuAspArgLysLysGlyTyrPheGlyValAsnLysLysHisIleMetAsp 430
Qy 746 -----ACCAACATACACAGTTGCATATGATGAATAAGTTCTGCTGGGATAAAG 793

Db 431 IleGlnAspThrAspGluAenThrGluProHisThrProPheGln-----Lys 446
Qy 794 CATGATTCTTTGAGAGGATGCACAAAACTTTGGCTGTTTCTGGAAGTGTCTTTTCTTTA 853
Db 447 TrpAspPheThrSerLysArgLysTyrPhePheillePheSerSerAlaValThrValAla 466
Qy 854 GGTGCGCTTGTCTCGGGTTTGGAGCTCGGAATTCCTGTTGGGAATGGATTTTAAAGGA 913
Db 467 GlyIleilleleLeuLeuValPheArgLeuAen-----LeuGlylleAspPheAlaSer 484
Qy 914 GGG-----TATGCGCTTTTACCTTTTAAATCCAAAAGACATGGC 949
Db 485 GlyAlaArgIleGluValGlnSerAspHisLysLeuThrThrGluGlnValGluLysAsp 504
Qy 950 ATCAGCGATGTTGTCAAAATCGCTGGCAAAATTTGTGCTAATAACTACAGGAAGCTGGTCTT 1009
Db 505 PheGluSerLeuGlyMetAspProAspThrValValLeuSerGlyGlyLysSerAsnIle 524
Qy 1010 TCT---TCTAGAGACTTCCGTATTCAAACTATTTGGATCTTCAGAAAAGATCAAAATCTAT 1066
Db 525 GlyValAlaArgPheValGlyValProAspLysGluThrIleAlaLysValLysThrTyr 544
Qy 1067 TTTAGTGTAAAGCTTTAAGCTATACTAAGCAGATACGAGCCTCTCTCTAAAATTAACG 1126
Db 545 PheLysAspLys----- 548
Qy 1127 ATCATGAGCTGGCGTTATTGTGGGATTGTTGTCAGAAAAGCGCTAGATTTCTCTACGGA 1186
Db 549 -----TyrGly 550
Qy 1187 AACTCTAAACGAAACGCAAAATTTTGTGTCAAAGTAAAGCAGCAAACTATCGAAGAAATG 1246
Db 551 SerAspProAsnValSerThrValSerProThrValGlyLysGluLeuAlaArgAsnAla 570
Qy 1247 CGTTATCAGCGCACCACCGGCTTTTAGGAGCTTTGGCAATCATCTTCCTCTATGTAGT 1306
Db 571 LeuTyrAlaValAlaIleAlaSerilleGly-----IleIleIleTyrValSer 586
Qy 1307 TTGGCTTTGAATGGCAATATGCTTTCAGTCCGCTATGCGCTTTAATTCATGACCTTTTG 1366
Db 587 IleArgPheGluTyrLysMetAlaIleAlaIleAlaIleAlaSerLeuLeuTyrAsp----- 604
Qy 1367 GCTACCTGTGCGCTTTGTTTATAGCACATTTCTTTTGAAGAAAAATTCAAATAGATTG 1426
Db 605 -----AlaPhePheIleValThrPhePheSerilleThrArgLeuGluValAspVal 621
Qy 1427 CAAGCCATTGCTGTTTAAAGCTGTTATGGGTATTCATTAACAATACATCTTTCATCATT 1486
Db 622 ThrPheilleAlaIleLeuThrIleIleGlyTyrSerilleAsnAspThrIleValThr 641
Qy 1487 TTTGATCGTATTCCTGAA-----GATCGCAAGCGAACCTGTTTACCCTTATGCAT 1537
Db 642 PheAspArgValArgGluHisMetLysLysArgLysProLysThrPheAlaAspLeuAsn 661
Qy 1538 GTTTTAGTAAATGATGCTTCAAAAGACGTTCAAGCCGACGGTAATACACAGCTACA 1597
Db 662 HisIleValAsnLeuSerLeuGlnGlnThrPheThrArgSerilleAsnThrValLeuThr 681
Qy 1598 ACTCTACAGTTTGTGTTTATAGCGGCTTGTATAGCGGCTCTCTGCTCTTAAATTTGCA 1657
Db 682 ValValIleValValValThrLeuLeuIlePheGlyAlaSerSerilleThrAsnPheSer 701
Qy 1658 TTTATTATGACATAGGATTTCTTAGGAATTTATCGTCTCTTTTATTTATGACCACCT 1717
Db 702 IleAlaLeuValGlyLeuLeuThrGlyValTyrSerSerLeuTyrIleAlaAlaGln 721
Qy 1718 CTGTTGTTG 1726
Db 722 IleTrpLeu 724
RESULT 3
US-09-881-752A-96

```

; Sequence 96, Application US/09881752A
; Patent No. US20020115078A1
; GENERAL INFORMATION:
; APPLICANT: Kleantchous, Harold
; APPLICANT: Al-Garawi, Areal
; APPLICANT: Miller, Charles
; APPLICANT: Tomb, Jean-Francois
; APPLICANT: Oomen, Raymond P.
; TITLE OF INVENTION: Identification of Polynucleotides
; TITLE OF INVENTION: Encoding No. US20020115078A1 Helicobacter Polypeptides in the
; FILE REFERENCE: 06132/041002
; CURRENT APPLICATION NUMBER: US/09/881,752A
; CURRENT FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: US 08/833,457
; PRIOR FILING DATE: 1997-04-01
; NUMBER OF SEQ ID NOS: 370
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 96
; LENGTH: 525
; TYPE: PRT
; ORGANISM: Helicobacter pylori
;
US-09-881-752A-96

Alignment Scores:
Pred. No.: 1.2e-35 Length: 525
Score: 431.00 Matches: 99
Percent Similarity: 60.83% Conservative: 47
Best Local Similarity: 41.25% Mismatches: 92
Query Match: 13.01% Indels: 2
DB: 10 Gaps: 2

US-09-868-987-1 (1-1864) x US-09-881-752A-96 (1-525)
QY 47 GGACATATTTCT--GCAACCGTGATGCGCTATGCTGTAGTATGACGTTATATG 103
Db 274 G|YAspPheserG|YAlaAsnValG|YAsrG|MetAlaIleValIleuAspAsnVal 293
QY 104 GTCAGCAGCCCTATTTAAACGTCCTCATG--AAAAATCATGCCAGTGTCTCAGGAAA 160
Db 294 TyserAlaProValIleArgIleuArgIleuGlyGlySerGlyGlnIleSerGlyAsn 313
QY 161 TTTACCAACCGTGAAGTGAAGCAATCGCTCAGTTTAAATGTGAGCGATCTTTT 220
Db 314 PheserValAlaGlnAlaSerAspLeuAlaIleAlaLeuAsrGlyAlaMetSerAla 333
QY 221 GTTCCCGAGTTCCTCAGTGAAGACGATCTTCTGTATCTTGGGAAAAAACAATGACA 280
Db 334 ProIleGlnValIleuGlnuLyAsrGleIleGlyProSerIleuGlyAspSerVallys 353
QY 281 CAAGCATTTATCTCAGCATGCTGTGGCTTGSCAATGCTTATGTTTGATGAGCGTATAT 340
Db 354 ThiserIleIleAlaLeuValIleGlyGlyPheIleuValMetGlyPheMetValIleuTy 373
QY 341 TATGATTTGAGCGCTCATGCTTGGAGCTGTCTCTGATCTTTTCTTCTTCTGATCTCG 400
Db 374 TyserMetAlaGlyValIleAlaCysLeuAlaLeuValAlaAsnLeuPheLeuIleVal 393
QY 401 GCACGTCTACAGTATTTGATGCGCACCTTGCTCAGACCTGCTGGAGATTGTTCTT 460
Db 394 AlaValMetAlaIlePheGlyAlaThreLeuThreProGlyMetAlaGlyIleValIleu 413
QY 461 GCTATGGGATGCGCGTATATGATGCAAAATGTTCTTATTCGAAGAATCGAGCAATTT 520
Db 414 ThrValGlyIleAlaValAspAlaAsnIleIleIleAsnGlnuArgIleArgGlnValIleu 433
QY 521 TTATGTCTCAAGCTTTAAAAATCTGTAGAAAAAGATATACCAAGCTTTTGAGACC 580
Db 434 ArgGlnuSngIuGlyIleAlaLysAlaIleHisLeuGlyTyrlAsnAlaSerArgAla 453
QY 581 ATTTTGTATCTAAGTACTGATGATTTGGCTCAGACCTCTTTTCTTCTTCTGATGACA 640
Db 454 IlePheAspSerAlaIleThreIleuIleAlaSerValIleuLeuTyrlAlaTyGlyThr 473

```

```

QY 641 GGGCTATTAAAGGTTTGCTTGAATGATTGATTTAGAAATTTCTCTCAATGTTAG 700
Db 474 G|YAlaIleIleuGlyPheAlaLeuThrGlnGlyIleGlyIleuAlaSerIleIleThr 493
QY 701 GCTCTTTTCATGACTTAAATTTTCTTCATGCTGTGATGATTAAGCCCAACATACAG 760
Db 494 AlaIleValGlyThrGlnGlyIleTyrglnAlaLeuLeuProLysLeuThrGlnThrLys 513

RESULT 4
US-09-899-482-4
; Sequence 4, Application US/09899482
; Patent No. US20020006641A1
; GENERAL INFORMATION:
; APPLICANT: Quax, Wilhelmus J.
; TITLE OF INVENTION: Increasing Production of Proteins in Microorganisms
; FILE REFERENCE: GC385-PCT
; CURRENT APPLICATION NUMBER: US/09/899,482
; CURRENT FILING DATE: 2001-07-05
; PRIOR APPLICATION NUMBER: EP 97305286.3
; PRIOR FILING DATE: 1997-07-16
; PRIOR APPLICATION NUMBER: EP 97305344.0
; PRIOR FILING DATE: 1997-07-17
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 323
; TYPE: PRT
; ORGANISM: E. coli
;
US-09-899-482-4

Alignment Scores:
Pred. No.: 6.47e-29 Length: 323
Score: 364.00 Matches: 102
Percent Similarity: 49.52% Conservative: 53
Best Local Similarity: 32.59% Mismatches: 124
Query Match: 10.99% Indels: 34
DB: 10 Gaps: 9

US-09-868-987-1 (1-1864) x US-09-899-482-4 (1-323)
QY 794 CATGATTTCTTGAGAGATGCAAAAACTTGGGCTGTTCTGGAAGTGTCTTTCTTTA 853
Db 17 TyAspPheMetArgTrpAspLyrrTrpAlaPheGlyIleSerGlyLeuLeuIleAla 36
QY 854 GGTTCGCTTCCTCGGCTTGGAGCGTGAATTCGTTTGGAAATGATTTTAAAGA 913
Db 37 AlaIleValIleMetGlyValArgGlyPheAsn-----TrpGlyLeuAspPheThrGly 54
QY 914 GGGTATGCCCTTACCTTTATCCAAAAGACATGCAATGATGCTGCTCAAAATGCC 973
Db 55 GlyThrValIleGluIleThreLeuGlnuLyProAlaGlnIleAspVal-----MetArg 72
QY 974 GGCAAAGTTGTGATTAACATACAGAGCTGCTTCTTCTGAGACCTTCGATTTCAA 1033
Db 73 -----AspAlaLeuGlnuLyAlaGlyPhe-----GluGlnProMetLeuGln 86
QY 1034 ACATTTGATCTTCAGAAAAGATCAAAATCTATTTAGATTAAGCTTTAAGCTTACT 1093
Db 87 AsnPheGlySerSerHisAspIleMetValArgMetProAlaGlnuGlyIuThrGly 106
QY 1094 AAGCATATGACCTCT--CTCTTAAATTTAAGCATATGAGCTTGCGCTTATTTGGG 1150
Db 107 GlyGlnValIleuGlySerGlnValIleuLyValIleAsnGlnuSer----- 121
QY 1151 ATGTGTGTCAAAAACAGGCTTAGTTCCTCTACGAAACTCTTAAACGAAACGAAATTT 1210
Db 122 -----ThrsGlnAsnAlaValIlyAsr 130
QY 1211 TGGTCAAGGTAAAGCAGCAAACTATCGAAGAAATCGTTATCAGGCGACATCGGCGTT 1270
Db 131 IleGlnPheValGlyProSerValGlyAlaAspLeuValGlnThrGlyAlaMetAlaLeu 150

```

```
QY 1271 TTAGGAGCTTTGGCAATCATCTTGCTCTATGTGAGTTTGGCTTTGAATGCAATATGCT 1330
Db 151 MetAlaLeuLeuSerIleLeuValTyrValGlyPheArgPheGluTrpArgLeuAla 170
QY 1331 TTCAGTGGCGGTATGGCTTTAATTCATGACCTTTGGCTTACCTGGCAGTCTTGTGTATA 1390
Db 171 AlaGlyValValIleAlaLeuAlaHisAspValIleIleThrLeuGlyIleLeuSerLeu 190
QY 1391 GCACATTTCTTTTGAAGAAATTCAAATAGATTGCAAGCATTGGTGCTTTAATGACT 1450
Db 191 PheHis-----IleGluIleAspLeuThrIleValAlaSerLeuMetSer 205
QY 1451 GTATTGGGATTCATTAAACAATCTTTGATCATTTTGTGATTCGTTATTCGCAAGAT--- 1507
Db 206 ValIleGlyTyrSerLeuAsnAspSerIleValValSerAspArgIleArgGluAsnPhe 225
QY 1508 GCCCAAGCGAACCTTTTACCCCTATGCATGTTTGTAGTTAATGATGCCCTTCMAAGACG 1567
Db 226 ArgGlyIleArgGlyThrProTyrGluIlePhe---AsnValSerLeuThrGlnThr 244
QY 1568 TTCAGCCCGCAGGTAAATGACACAGCTACAATCTATCAGTTTGTGTTTGTGTTT 1627
Db 245 LeuHisArgThrLeuIleThrSerGlyThrLeuMetValIleLeuMetLeuTyrLeu 264
QY 1628 ATAGCGGCTCTCTGCTCTTTAATTTTCATTTATATGACCATAGGATTTCTTACGA 1687
Db 265 PheGlyGlyProValLeuGluGlyPheSerLeuThrMetLeuIleGlyValSerIleGly 284
QY 1688 ACTTATCTCTCTTTATATGACCACTCTGTGTGTTG 1726
Db 285 ThrAlaSerSerIleTyrValAlaSerAlaLeuAlaLeu 297
RESULT 5
US-09-899-482-3
; Sequence 3, Application US/09899482
; Patent No. US20020006641A1
; GENERAL INFORMATION:
; APPLICANT: Quax, Wilhelmus J.
; TITLE OF INVENTION: Increasing Production of Proteins in Microorganisms
; FILE REFERENCE: GC385-PCT
; CURRENT APPLICATION NUMBER: US/09/899,482
; CURRENT FILING DATE: 2001-07-05
; PRIOR FILING DATE: 1997-07-16
; PRIOR APPLICATION NUMBER: EP 97305286.3
; PRIOR FILING DATE: 1997-07-17
; PRIOR APPLICATION NUMBER: EP 97305344.0
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 615
; TYPE: PRT
; ORGANISM: E. coli
US-09-899-482-3
Alignment Scores:
Pred. No.: 2.44e-28 Length: 615
Score: 359.50 Matches: 80
Percent Similarity: 62.39% Conservative: 61
Best Local Similarity: 35.40% Mismatches: 82
Query Match: 10.85% Indels: 3
DB: 10 Gaps: 2
US-09-868-987-1 (1-1864) x US-09-899-482-3 (1-615)
QY 44 AATGGACAATATCTTCAACCGTGTAGTGGCTTAGTCAATGACGGTTATATG 103
Db 375 SerGlyLysLysAspAlaAsn-----GlyArgAlaValLeuValIleGlnGluVal 392
QY 104 GTCAGCAGCCCTATTTTAAACGCTCCCATGAAATCATGCCAGTGTCTCAGGAGAAATTT 163
Db 393 IleAsnIleAlaAsnIleGlnSerArgLeuGlyAsnSerPheArgIleThrGlyIleAsn 412
QY 164 ACCCAACCGTGAAGTAGGCAAACTCGCCTCAGATTTTAAATCTGGAGCGCATGTCTTTGTT 223
```

```
Db 413 AsnProAsnGluAlaArgGlnLeuSerLeuLeuLeuArgAlaGlyAlaLeuIleAlaPro 432
QY 224 CCCGAGGTTCCTCAGTGAAGAGACGATCTCTTCTCATCTTGGGAAAAACAATGTACACAA 283
Db 433 IleGlnIleValGluAlaArgThrIleGlyProThrLeuGlyMetGlnAsnIleGluGln 452
QY 284 GGCATATCTCAGCATGCTGTGGCTTGGCAATGCTTATTTGTTGATGAGCGTATATAT 343
Db 453 GlyLeuGluAlaCysLeuAlaGlyLeuValSerIleLeuPheMetIleIlePheTyr 472
QY 344 AGATTTGGAGCGCATCGCTTGGGAGCTGTTCTCTCTGAATCTTTGCTTATCTGGCA 403
Db 473 LysLysPheGlyLeuIleAlaThrSerAlaLeuIleAlaAsnLeuIleValGly 492
QY 404 GCTCTACAGTATTG---GATGCCCACTCACCTTGTTCAGGACTCGCTGGGATTTGTTCTT 460
Db 493 IleMetSerLeuLeuProGlyAlaThrLeuSerMetProGlyIleAlaGlyIleValLeu 512
QY 461 GCTATGGGATGCGCGTAGATGCAATGTTCTTGTATTCGAAAGAAATCCGAGAGAAATTT 520
Db 513 ThrLeuAlaValAlaValAspAlaAsnValLeuIleAsnGluArgIleLysGluGluLeu 532
QY 521 TTATTGCTCAAGTCTTAAAAATCTGTAGAAAAGATATACCAAGCTTTTGGAGCC 580
Db 533 SerAsnGlyArgThrValGlnGlnAlaIleAspGluGlyTyrArgGlyAlaPheSerSer 552
QY 581 ATTTTTCATCTTAATCTGACTACAGTATTGGCCCTCAGCACTTCTTTCTCTCTAGATACA 640
Db 553 IlePheAspAlaAsnIleThrThrLeuIleValIleIleLeuTyrAlaValGlyThr 572
QY 641 GGCCTATTAAAGGTTTGTCTTGCATGATGATTGAGAAATTTCTCTCAATGTTTACG 700
Db 573 GlyAlaIleLysGlyPheAlaIleThrThrGlyIleGlyValAlaThrSerMetPheThr 592
QY 701 GCTCTTTTCATGACTAAA 718
Db 593 AlaIleValGlyThrArg 598
RESULT 6
US-09-852-053-3
; Sequence 3, Application US/09852053
; Patent No. US20020055141A1
; GENERAL INFORMATION:
; APPLICANT: BERENS, STEPHAN
; APPLICANT: KALINOWSKI, JORN
; APPLICANT: PUHLER, ALFRED
; TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM STRAIN WITH
; ENHANCED SECRETION ACTIVITY
; FILE REFERENCE: WAS/21123/280248
; CURRENT APPLICATION NUMBER: US/09/852,053
; CURRENT FILING DATE: 2001-05-10
; PRIOR APPLICATION NUMBER: EPO 00110021.3
; PRIOR FILING DATE: 2000-05-12
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 637
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
; FEATURE:
; NAME/KEY: PROPEP
; LOCATION: (1)..(637)
; OTHER INFORMATION: secD
US-09-852-053-3
Alignment Scores:
Pred. No.: 5.5e-20 Length: 637
Score: 277.50 Matches: 67
Percent Similarity: 51.81% Conservative: 62
Best Local Similarity: 26.91% Mismatches: 87
Query Match: 8.38% Indels: 33
DB: 10 Gaps: 7
```



```
Db 256 eGlyThrTyrSerSerLeuPheIleAlaThrArgIleTrpIleIleLeuGluSerSerAr 276
Qy 1744 AAATGCG 1750
Db 276 gAsnArg 278
RESULT 8
US-09-852-053-4
; Sequence 4, Application US/09852053
; Patent No. US20020055141A1
; GENERAL INFORMATION:
; APPLICANT: BERENS, STEPHAN
; APPLICANT: KALINOWSKI, JORN
; APPLICANT: PUHLER, ALFRED
; TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM STRAIN WITH
; FILE REFERENCE: MAS/21123/280248
; CURRENT APPLICATION NUMBER: US/09/852,053
; CURRENT FILING DATE: 2001-05-10
; PRIOR APPLICATION NUMBER: EPO 00110021.3
; PRIOR FILING DATE: 2000-05-12
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 403
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
; FEATURE:
; NAME/KEY: PROPEP
; LOCATION: (1)..(403)
; OTHER INFORMATION: secf
US-09-852-053-4
Alignment Scores:
Pred. No.: 4,95e-15 Length: 403
Score: 228.00 Matches: 72
Percent Similarity: 44.97% Conservative: 80
Best Local Similarity: 21.30% Mismatches: 132
Query Match: 6.88% Indels: 54
DB: 10 Gaps: 8
US-09-868-987-1 (1-1864) x US-09-852-053-4 (1-403)
Qy 797 GATTCTTGTAGAGATGCAAAACCTTTGGCTGTTCCTGGAAGTGTTTTCTTTTAGGT 856
Db 34 AspPheIleAlaLysThrLysLeuTrpTyrTrpIleThrGlyIleLeuValIleSer 53
Qy 857 TCGGTGCTCTCGGTTTGGAGCCCTGGAATCCGTTTGGGAATGGATTAAAGGAGGG 916
Db 54 IleLeuPheIleAlaLysThrLysLeuTrpTyrTrpIleThrGlyIleLeuValIleSer 71
Qy 917 TATGCTTTTACCTTTTAAATCCAAAGAGCATGGCATCGCATGTTGCTCAATGCGTGGC 976
Db 72 ThrLysMetSerMetProAlaSerAspTyrSerThrGluGlnValGluGluThr----- 89
Qy 977 AAAGTTGTGCATAAATACTACAGAAAGCTGCTTCTTCTAGAGACTTCGCTATTCAACA 1036
Db 90 -----PheThrGluAlaThrGlyIleThrProGluIleValGlnIleValGly 105
Qy 1037 TTTGGATCTTCAGAAAGATCAAAATCTATTTTAGTGAATGCTTTAAAGCTTACTACTAAG 1096
Db 106 SerGlyAspAlaArgThrLeuGluIleTyr-----SerGluArgLeuSerAspGluAsp 123
Qy 1097 CAGATACAGAGCTCTCTCTCTAAATAACGATCATGAGCTGCGTTATTGTGGGATGTT 1156
Db 124 ValGluLysAlaArgLeu-----AlaIleTyr 132
Qy 1157 GTCGAAACAGCCCTAGATTTCTCTACGAAACCTCAACGAAACGCAAAATTTTGTGCA 1216
Db 133 GluGluTyrGlnProLeuAsnSerGluGlnProSerProAspAlaIleGlyAsnSer 152
Qy 1217 AAGGTAACGACGCAAACTATCGAGAAATGCGTTATCAGGGACCATCGGCTTTTAGGA 1276
Db 1217 AAGGTAACGACGCAAACTATCGAGAAATGCGTTATCAGGGACCATCGGCTTTTAGGA 1276
```

```
Db 153 ThrValSerGluSerTrpGlySerThrIleThrGlnArgMetValLeuAlaLeuIleAla 172
Qy 1277 GCTTTGGCAATCATCTTCTCTATGTAGTTTGGCTTTTGAATGGCAATATGCTTTTCACT 1336
Db 173 PheLeuValIleAlaAlaIleTyrIleAlaPheArgLeuGluArgGluMetAlaIleAla 192
Qy 1337 GCGTATGCGCTTTAATTCATGACCTTTTGGCTACCTGTGTGCAGCTCTGTTTATAGCACAT 1396
Db 193 AlaMetAlaAlaLeuValValAsp-----GlyIleValIleAlaGlyIle 207
Qy 1397 TCTTTTGGAGAAAATTCAAATAGATTGTGCAAGCCATTGTGCTTTTAATCACTGTATTG 1456
Db 208 TyrAlaValIleGlyLeuGluValSerProAlaThrValIleGlyLeuLeuThrValLeu 227
Qy 1457 GGGTATTCAATTAACAATCTTTCATCATTTTTCATCGTATTTCGTGAAGAT----- 1507
Db 228 ThrPheSerIleTyrAspThrValValValPheAspLysValArgGluAsnThrGluGly 247
Qy 1508 -----CGCCAAGGGAACCTGTTTACCCCTATGTCAT 1537
Db 248 PheGluGlySerArgArgThrTyrAlaGluGlnAlaAsnLeu----- 262
Qy 1538 GTTTTAGTTAATGATGCTTCAAAAGACGTTTCAGCGCAGGTAATGACACACAGCTACA 1597
Db 263 -----AlaValAsnGlnThrPheMetArgSerIleSerThrIleIle 277
Qy 1598 ACTCTATCAGTTTGTGTTAATGCTTTTGTATATA-----GGCGGCTCTCTCT 1642
Db 278 SerAlaLeuProIleIleAlaLeuMetValValAlaValTyrMetMetGlyValGlyThr 297
Qy 1643 GTCCTTTAATTTTGGATTTTATGACCATGAGGATTTCTTCTAGGAATTTATGCTCTCTT 1702
Db 298 LeuLysAspLeuAlaLeuIleGlnLeuIleGlyValIleGluGlyThrPheSerSerVal 317
Qy 1703 TATATTGCACCATCTGCTGCTGCTTTTATGCTCGTAAAGAAAATCGCTCAAAA 1756
Db 318 PheLeuAlaThrProLeuLeuValSerLeuLysAsnArgLeuSerLysThrLys 335
RESULT 9
US-09-815-242-12189
; Sequence 12189, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12189
```

```

; LENGTH: 1055
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-12189

Alignment Scores:
  pred. No.:      1.16e-06      length:      1055
Score:           147.50        Matches:      132
Percent Similarity: 34.42%      Conservative: 111
Best Local Similarity: 18.70%      Mismatches: 229
Query Match:      4.45%         Indels:      234
DB:               10           Gaps:      28

US-09-868-987-1 (1-1864) x US-09-815-242-12189 (1-1055)

QY 265 GAAAAAACAATGTACACAGGCAATTATCTACGATGCTGGCTTGGCAATGCTTATGT 324
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 366 GluysSerleuYrThrMet-ValGluYAlaSerleuGlYrThrleValAlaIle1 385
QY 325 TTTCATGAGCGTATATTAGA-----TTTGAGGCGTCAATGCTTCGGAGCTGTCT 378
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 385 eValIleleuYrThrMetValGluYAlaSerleuGlYrThrleValAlaIle1Pr 405
QY 379 TCTGAATCTTTGCTTATCTGAGCGCTCTACAGATTTGGATGCGCACTGCTTGTG 438
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 405 oleuSerleuYrThrMetValGluYAlaSerleuGlYrThrleValAlaIle1Pr 425
QY 439 AGGACGCTGGGATGCTTCTGCTATGGGATGCGCGTATGCAATGCTTGTAT 498
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 425 uThrleuGlYAlaSerleuValAlaIleGlYArgValIleAspSerleValAla 445
QY 499 CGAAAGATC-----CGAGAGATTTTATTGCTCAAACTCTTAA 540
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 445 lGluAsnIleYrArgIleuThrAspSerleuGlYrThrleValAlaIle1Pr 464
QY 541 AAATCTGTGAAAGATATACCAAGGCTTTTGAGCCCTTTTGATCTTAATGAC 600
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 465 -----IleIleSerIleThrGluValIleuYrProIleuSerleuThrleuVa 482
QY 601 TACAGTATGGCTCAGCACTTCTTCTCTCTATGATACAGGCGCTAT 649
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 482 lThrIleIleValIleuYrProleuValIleValIleSerIleYserValIleuMetPh 501
QY 650 -AAAGGCTTCTTGAATGATTTAGCAATTTCTCTTCAATGTTTACG----- 700
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 501 eArgProIleuAlaIleuAlaIleuAlaIleuAlaIleuAlaIleuAlaIleuAlaIleu 521
QY 701 -----GCTCTTTCATGACTAAATTTTCTTCATGCTGAGATGATAGACCA 750
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 521 rLeuValIleuAlaIleuAlaIleuAlaIleuAlaIleuAlaIleuAlaIleuAlaIleu 541
QY 751 ACATACACAG-----TTGCATATGATG- 772
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 541 nHISGInGluYrleuGlYAlaValSerThrThrIleYrIleYrValIleuHISrIle 561
QY 772 ----- 772
Db 561 uAsnHISrIleValIleIleIleuSerThrIleuIleuValAlaThrIleValAlaPh 581
QY 773 -----AATAAGTTCGGGGAT 789
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 581 eGlYrProIleuGlYrThrSerPheIleSerAlaGlyAspAspIleuAlaIle 601
QY 790 AAACAT----- 796
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 601 eThrTYrThrProIleuYrGluYrThrGluGlnAlaValAlaIleuAsnHISAlaGlnAspVa 621
QY 797 -----GATTTCTGAGGATGCAAAAATTGGGCTGT----- 832
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 621 lGluYrleuYrGlnIleYrSHISValIleYrThrIleGlnTYrSerValIleYrGlySe 641
QY 833 -----TCGGAAGTGTCTTTTCTTTAGGTTCGCTTGT----- 865
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::

```

```

Db 641 rSerProValAspProThrGlySerThrAsnSerMetAlaIleMetValGluTYrAspAs 661
QY 865 ----- 865
Db 661 nAsrThrProAsnPheAspValGluAlaAspValIleYrSHISAlaAspGlyPheY 681
QY 866 -----CTCGGGTTTGGAGCTGGAATTCGTTTGGG 897
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 681 SHISrProGluYrTrPLeAsnGlnAspIleuYrGluYrAlaGlyAsnYserValG 701
QY 898 AATGATTTTAAAGAGGATGCTTTCATCTTAAATCCAAAAGCATGCGATCGACGA 957
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 701 uValThrValIleYrGlyPro-----SerMetAspAlaIleYrSerThrValYsAs 728
QY 958 TGTTCCTCAATGCTGCGCAAGTT----- 982
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 718 pIleGluGlnYrMetIleYrGlnValIleYrGlyLeuAlaAsnValYrSerAspSerleu 738
QY 983 -----GTGCAATTAACCTACAGGAAGCTGCTTTC 1011
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 738 nHrTYrAspGlnTYrGluIleYrValAspGlnAsnYrAlaAlaGluAsnGlyIleSe 758
QY 1012 TTCTAGAGACTTCCTGAT 1030
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 758 rAlaSerGlnLeuValMetHISleuAsnGluAsnLeuProGluYrThrValIleThrVa 778
QY 1031 -----CAACATTTGGATC 1044
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 778 lYrGluAsnGlyLeYrThrValGlyValIleYrValYrGlnAsnYrGlnThrAspTrpSe 798
QY 1045 TTCAAGAAAGATCAAAATGATTTTGTATAGCTTAAAGCTTAAAGATACAGAGATAC 1104
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 798 rGluAspYrLeuAsnAsnIleThrleuYrSlyrProThrGlyYrThrIleYrleuGl 818
QY 1105 A-----GCCTCTCTCTAAATTAACGATCATGAC-----TGGCG 1140
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 818 yAspIleAlaIleThrleuValYrThrThrProserIleYrleuThrGlnGluGlnYrAs 838
QY 1141 TTAATTGGGATGTT-----GTCAAGAACAGCGCTAGATTCTCTACGAAATC 1191
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 838 pTYrAlaThrThrValSerAlaYrValIleThrleuYrAsp-----ValGlyGlyTh 855
QY 1192 TAAACGAAGCAAAATTTGTCAAGGTACACCAAACTGACGAAGAAATGGCTTA 1251
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 855 rHnArg-----GlnAlaMetSerIleAsnAsn-----LeuAspYrProAsnAsnVa 872
QY 1252 TCAGGCGACCATCGCG----- 1267
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 872 lYrSValAsnIleGlyYrAlaSerAspAspIleAsnAsnAlaMetThrGlnleuAlaPh 892
QY 1268 -----CTTTAGAGACTTTGGCAATCATCTTGCTATGATGAGTTGCGCTTGAATGCA 1323
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 892 eAlaMetleuAlaIleIleIleValIleYrleuIleleuValIleThrPheYsGlyG 912
QY 1324 ATATGCT-----TTCAGTCCGCTATGCGCTTAATTCATGACCTTTGGCTACGCTGAGCT 1380
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 912 yLeuAlaProPheThrIleuPheSerleuProPheThrValIleGlyValIleIleAl 932
QY 1381 CTGTTTATAGACATTTCTTTTGAAGAAATTCAAATGATTTGCCAAGCCATGGTGTC 1440
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 932 aLeuIleIleThrGly-----GluThrIleSerValPro-----SerleuIleGlyMe 948
QY 1441 TTTATGATCTGATTTGGGATTCATTTAAACAATCTTGATCATTTTGGATGATTCG 1500
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 948 tLeuMet-----LeuIleGlyIleValIleThrAsnAlaIleValIleuIleAspYrValI 967
QY 1501 TGAAGATCGCCAGACCAACCTGTTTACCCCTATGATGTTTATGTAATGATGCCCTTCA 1560
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 967 eAsnAsnGlnGlnGlnGlyMet-----GluMetYrGlnAlaLeuIleGluAlaGly 985
QY 1561 AAAGACGCTCAGCGGCAAGCTTAATGACACAGCTTAACCTATGATTTTGTAAATGCT 1620
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 985 YrThrArgIleArgProIleuMetThrAlaIleAlaThrIleGlyYrAlaLeuValProle 1005
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::

```

```

Qy 1621 TTTGTTATAGCGCTCTCTGCTCTT-----AATTTTGCAATTATTATGACCAT 1671
||||| ||||| ||||| ||||| |||||
Db 1005 uLeuPheGlyGlnAspSerSerIleLeuIleSerIleGlyLeuAlaAlaThrValIleG1 1025
||||| ||||| ||||| ||||| |||||
Qy 1672 AGGATTCTTCTAGGAATTTATGCTCTCTTATATGACCATCTCG-----TTGTT 1725
||||| ||||| ||||| ||||| |||||
Db 1025 yGlyLeuIleSerSerThrLeuLeuThrLeuValValProValIleTyrGluIleLe 1045
||||| ||||| ||||| ||||| |||||
Qy 1726 GTTATGTCCTGATAA 1741
||||| ||||| ||||| ||||| |||||
Db 1045 uPheThrLeuLysLys 1050
||||| ||||| ||||| ||||| |||||

RESULT 10
US-09-815-242-10331
; Sequence 10331, Application US/09815242
; Patent No. US2002061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR APPLICATION NUMBER: 2001-03-21
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10331
; LENGTH: 1034
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-815-242-10331

Alignment Scores:
Pred. No.: 2,62e-06 Length: 1034
Score: 144.00 Matches: 135
Percent Similarity: 31.73% Conservative: 89
Best Local Similarity: 19.12% Mismatches: 226
Query Match: 4.35% Indels: 256
DB: 10 Gaps: 30

US-09-868-987-1 (1-1864) x US-09-815-242-10331 (1-1034)
Qy 308 TTGGCAATGCTTATTTGTTGATGAGCGTATATTATAGATTTGGAGGCGTCATGCTTCG 367
||||| ||||| ||||| ||||| |||||
Db 353 LeuValMetTyrLeuPheLeuGlnAsnMetArgAlaThrLeuIleProThrIleAlaVal 372
||||| ||||| ||||| ||||| |||||
Qy 368 GGAGCTGTTCTCTGAATCTTTTCTATCTGGCAGCTCTACAGTATTTGGATGCGCCA 427
||||| ||||| ||||| ||||| |||||
Db 373 ProValValLeuLeuGlyThrPheAlaIleLeuAlaAlaPheGlyTyrSerIleAsnThr 392
||||| ||||| ||||| ||||| |||||
Qy 428 CTCACCTGTGAGGCTCGCTGGGATTTCTTCTGCTATGGGATGGCGGTAGATGCAAAAT 487
||||| ||||| ||||| ||||| |||||

```

```

Db 393 LeuThrMetPheGlyMet-----ValLeuAlaIleGlyLeuLeuValAspAspAla 409
Qy 488 GTTCTGTTATTCGAAAGAAATCCGAGAGAAATTTTATTGTCTCAAGTCTT-----AAA 541
||||| ||||| ||||| ||||| |||||
Db 410 IleValValGluAsnVal---GluArgValMetMetGluAspLysLeuProLys 428
||||| ||||| ||||| ||||| |||||
Qy 542 AATCTGTAGAAAAGATATATCAAGCGCTTTTGAGCCATTTTGTGATTCTTAACCTTGACT 601
||||| ||||| ||||| ||||| |||||
Db 429 GluAlaThrGluLysSerMetSerGlnIleGlnGlyAlaLeuVal-----GlyIleAla 446
||||| ||||| ||||| ||||| |||||
Qy 602 ACAGTATTGGCTCAGCACCTTCTT-----TTCCTTCTAGATACAGCCCTTAT 649
||||| ||||| ||||| ||||| |||||
Db 447 MetValLeuSerAlaValPheIleProMetAlaPheGlyGlySerThrGlyAlaIle 466
||||| ||||| ||||| ||||| |||||
Qy 650 ---AAAGGGTTTGTCTTTCACATGATTATTAGGAATTTTCTTCAATGTTTACGGCTCTT 706
||||| ||||| ||||| ||||| |||||
Db 467 TyrArgGlnPheSerIleThrIleValSerAlaMetAlaLeuSerValLeuValAlaLeu 486
||||| ||||| ||||| ||||| |||||
Qy 707 TTCATGACT----- 715
||||| ||||| ||||| ||||| |||||
Db 487 IleLeuThrProAlaLeuCysAlaThrLeuLysProValSerAlaGluHisGlu 506
||||| ||||| ||||| ||||| |||||
Qy 716 ---AAATTTTCTTCATGCTGCTGATGAATAAGACC---CAACATACACACTTGCATATG 769
||||| ||||| ||||| ||||| |||||
Db 507 AsnLysGlyGlyPhePheGlyTyrPheAsnThrThrPheAspHisSerValAsnHisTyr 526
||||| ||||| ||||| ||||| |||||
Qy 770 ATGAAT-----AAAGTTCGCGGATAAAGCATGATTTCTTG----- 805
||||| ||||| ||||| ||||| |||||
Db 527 ThrAsnSerValGlyLysIleLeuGlySerThrGlyArgTyrLeuLeuIleTyrAlaLeu 546
||||| ||||| ||||| ||||| |||||
Qy 806 ---AGAGATCCAAAAAATTTTGGCGCTTCTCGAAGCTGTTTCTT----- 850
||||| ||||| ||||| ||||| |||||
Db 547 IleValAlaGlyMetValLeuPheLeuArgLeuProSerSerPheLeuProGluGlu 566
||||| ||||| ||||| ||||| |||||
Qy 851 ---TTAGTTGGCTGCTCTCGGTTTGGAGCTGGAAT----- 886
||||| ||||| ||||| ||||| |||||
Db 567 AspGlnGlyValPheLeuThrMetIleGlnLeuProAlaGlyAlaThrGlnGluArgThr 586
||||| ||||| ||||| ||||| |||||
Qy 886 ----- 886
||||| ||||| ||||| ||||| |||||
Db 587 GlnLysValLeuAspGlnValThrAspTyrTyrLeuLysAsnGluLysAlaAsnValGlu 606
||||| ||||| ||||| ||||| |||||
Qy 887 ---TCCGTTTGGGAATGGATTTTAAAGGA----- 913
||||| ||||| ||||| ||||| |||||
Db 607 SerValPheThrValAsnGlyPheSerPheSerGlyGlnAlaGlnAsnAlaGlyMetAla 626
||||| ||||| ||||| ||||| |||||
Qy 913 ----- 913
||||| ||||| ||||| ||||| |||||
Db 627 PheValSerLeuLysProTyrGluArgAsnGlyAspGluAsnSerAlaGluAlaVal 646
||||| ||||| ||||| ||||| |||||
Qy 914 -----GGGTATGCTTTTACCTTTTAAAT 934
||||| ||||| ||||| ||||| |||||
Db 647 IleHisArgAlaLysMetGluLeuGlyLysIleArgaspGlyPheValIleProPheAsn 666
||||| ||||| ||||| ||||| |||||
Qy 934 ----- 934
||||| ||||| ||||| ||||| |||||
Db 667 MetProAlaIleValGluLeuGlyThrAlaThrGlyPheAspPheGluLeuIleAspGln 686
||||| ||||| ||||| ||||| |||||
Qy 934 ----- 934
||||| ||||| ||||| ||||| |||||
Db 687 AlaGlyLeuGlyHisAspAlaLeuThrGlnAlaArgAsnGlnLeuLeuGlyMetAlaAla 706
||||| ||||| ||||| ||||| |||||
Qy 935 -----CCAAAAGATCGCATCGCATGCTTGTCTCAAAATG 970
||||| ||||| ||||| ||||| |||||
Db 707 GlnHisProAlaSerLeuValSerValArgProAsnGlyLeuGluAspThrAlaGlnPhe 726
||||| ||||| ||||| ||||| |||||
Qy 971 CGTGGCAAAGTT---GTGCATAAACTACAGGAAGCTGGTCTTCTTCTTAGAGACTTCGCT 1027
||||| ||||| ||||| ||||| |||||
Db 727 LysLeuGluValAspGlnGluLysAlaGlnAlaLeuGlyValSerLeuSerAspIleAsn 746
||||| ||||| ||||| ||||| |||||
Qy 1028 ATTCAACATTTGATCTTCA----- 1048
||||| ||||| ||||| ||||| |||||
Db 747 ---GlnThrIleSerThrAlaLeuGlyThrTyrValAsnAspPheIleAspArgGly 765
||||| ||||| ||||| ||||| |||||

```

QY 1049 -----GAAAGATCAAAATCTATTTTATGATGATAA 1078
 Db 766 ArgValIylsLysLeuYrValGlnAlaAspAlaLysPheArgMetLeuProGluAspVal 785
 QY 1079 GCTTAAGCTTACTAGACATACGAGCCCTCTGCTCA-----AAATTAACGATC 1129
 Db 786 AspIylsLeuYrValArgSerAlaAsnGlyGluMetValProPheSerAlaPheThrThr 805
 QY 1130 ATGAGCTG-----CGTATTTGGGATTT---GTTGCAGA 1162
 Db 806 SerHisTrpValYrGlySerProArgLeuGluArgTrpYrAsnGlyLeuProSerMetGlu 825
 QY 1163 AACAGGCTAGATTCTTCTACGAAACTCTTAACGAAACGA---AAATTTGGTCAAG 1219
 Db 826 IleGlnGlyGlnAlaAlaProGlyThrSerSerGlyAspAlaMetAlaLeuMetGluAsn 845
 QY 1220 GTAGCAGCAACTATGC-----AGAAATGCGTTATCAGCGC 1258
 Db 846 LeuAlaSerIylsLeuPheAlaGlyIleGlyTrpAspTrpThrGlyMetSerTrpGlnGlu 865
 QY 1259 ACCATCCGG-----CTTTAGAGGCTTTGGCAATCATCTCTCATGTG 1303
 Db 866 ArgLeuSerGlyAsnGlnAlaProAlaLeuValAlaIleSerPheValValPheLeu 885
 QY 1304 AGTTTGGCGCTTGAATGCAATATGCTTTCAGTCCGCTTATTCATGACCTT 1363
 Db 886 CysLeuAlaAlaLeuYrGlySerTrpSerIleProValSerValMetLeuValValPro 905
 QY 1364 TTGGCTACCTTGACAGTCTTTTATAGCACAATTTCTTTGAGAAATTCATATGAT 1423
 Db 906 LeuGlyIleValGlyValLeuLeuAlaAlaThrLeuPheAsnGlnLys-----AsnAsp 923
 QY 1424 TTGCAGCGCTTGTGCTTATGACTGATGATGGGATTCATTAACATATCTTGTATC 1483
 Db 924 ValIyrPheMetValGlyLeuLeuThrThrIleGlyLeuSerAlaLysAlaIleLeu 943
 QY 1484 ATTTTGTATGCTATTCGTGAAGATCGCAGACGAACTGTTTACCCTTATGCAATGTTTA 1543
 Db 944 IleValGlu-----Phe 947
 QY 1544 GTTATATGATGCCCTTCAAG-----ACGTTACGCGGAGGTATGACACA 1591
 Db 948 AlaLysAspLeuMetGluLysGlnGlyIylsGlyValValGlnAlaThrLeuMetAlaVal 967
 QY 1592 GCTCAACTATCATAGTTTGTATATG-----CTTTGTTTATA----- 1630
 Db 968 ArgMetArgLeuArgProIleLeuMetThrSerLeuAlaPheIleLeuGlyValLeuPro 987
 QY 1631 -----3GCGGCTCTCTGCTTTTAAATTTTGCATTTATATGACATA 1672
 Db 988 LeuAlaIleSerAsnGlyAlaGlySerGlyAlaGlnAsn-----AlaValGlyIle 1004
 QY 1673 GGGATTTCTTAGGAACCTTATCGTCTTATATATGCA-----CCACCTCTGTG 1723
 Db 1005 GlyValMetGlyGlyMetValSerAlaThrLeuLeuAlaIlePhePheValProValPhe 1024
 QY 1724 TTGTTATGTCCTGATAA 1741
 Db 1025 PheValValIleAlaArg 1030
 RESULT 11
 US-09-815-242-11145
 ; Sequence 11145, Application US/09815242
 ; Patent No. US20020061569A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Haseilbeck, Robert
 ; APPLICANT: Ohlsen, Karl
 ; APPLICANT: Zyskind, Judith W.
 ; APPLICANT: Wall, Daniel
 ; APPLICANT: Trawick, John D.
 ; APPLICANT: Carr, Grant J.
 ; APPLICANT: Yamamoto, Robert T.

; APPLICANT: Xu, H. Howard
 ; TITLE OF INVENTION: Identification of Essential Genes in
 ; FILE OF INVENTION: Prokaryotes
 ; FILE REFERENCE: ELITRA.011A
 ; CURRENT APPLICATION NUMBER: US/09/815,242
 ; PRIOR FILING DATE: 2001-03-21
 ; PRIOR APPLICATION NUMBER: 60/191,078
 ; PRIOR FILING DATE: 2000-03-21
 ; PRIOR APPLICATION NUMBER: 60/206,848
 ; PRIOR FILING DATE: 2000-05-23
 ; PRIOR APPLICATION NUMBER: 60/207,727
 ; PRIOR FILING DATE: 2000-05-26
 ; PRIOR APPLICATION NUMBER: 60/242,578
 ; PRIOR FILING DATE: 2000-10-23
 ; PRIOR APPLICATION NUMBER: 60/253,625
 ; PRIOR FILING DATE: 2000-11-27
 ; PRIOR APPLICATION NUMBER: 60/257,931
 ; PRIOR FILING DATE: 2000-12-22
 ; PRIOR APPLICATION NUMBER: 60/269,308
 ; PRIOR FILING DATE: 2001-02-16
 ; NUMBER OF SEQ ID NOS: 14110
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 11145
 ; LENGTH: 1032
 ; TYPE: PRF
 ; ORGANISM: Haemophilus influenzae
 ; US-09-815-242-11145
 Alignment Scores:
 Pred. No.: 1,52e-05 Length: 1032
 Score: 136.50 Matches: 136
 Percent Similarity: 34.918 Conservative: 145
 Best Local Similarity: 16.898 Mismatches: 241
 Query Match: 4.12% Indels: 23
 DB: 10 Gaps: 30
 US-09-868-987-1 (1-1864) x US-09-815-242-11145 (1-1032)
 QY 44 AATGACATATTTCTGAAACCGTAGTGCGTATGATGATGACGTTATATG 103
 Db 274 AsnSerArgAlaThrAlaAsnGlyAlaGlySerValValLeuAlaIleAsnProThrSer 293
 QY 104 GTACAGACGCCCTATTTTAAACGCCATTGAAAAATCATGCCAGTCTTCAGGAAATTT 163
 Db 294 ThrAlaAsnProLeu-----ThrValAlaGluLys--- 303
 QY 164 ACCACCGTGAAGTGAACAACCTCGCTCATTTTAAATCTGAGCGATG---TCTTTT 220
 Db 304 -----IleArgProLeuYrGlySerIleIleYrThrGlnLeuProAspSerMet 319
 QY 221 GTTCCGAGGTTCTCAGTGAAGAGCATCTCTTGATCTTGGGAAAAACAATGTACA 280
 Db 320 GluSerAspIleLeuYrAspArgThrIleAlaIleAsnSerIleHisGluValIle 339
 QY 281 CAAAGCATTTCTCAGATCCTGTGGCTTGGCATGTTATTTTGTATGAGACGTATAT 340
 Db 340 LysThrIle-----GlyGlnAlaThrLeuIleValLeuValIleLeu 354
 QY 341 TATAGATTTGAGAGC-----GTATCGCTTGGGAGCTGTCTTCTGAATCTT 388
 Db 355 MetPheIleGlySerPheArgAlaIleLeuIleProIleLeuAlaIleProIleSerLeu 374
 QY 389 TTGCTATCTGGGACGCTTACAGTATTTGGATGCGCACTTGTTCAGGATCGCT 448
 Db 375 IleGlyValLeuMetLeuLeuGlnSerPheAsnPheSerIleAsnLeuMetThrLeuLeu 394
 QY 449 GGGATTTCTTCTGATGGGATGCGCGTATGATGCAATCTTCTGTATTCGAAAGATC 508
 Db 395 AlaLeuIleLeuAlaIleGlyLeuValValAspAspAlaIleValValLeuGluAsnIle 414
 QY 509 CGAGAGAAATTTTATGTCTCAAGTCTTAAAAAATCTGTAGAAAAAGATATACCAAG 568
 Db 415 -----AspArgHisIleLysAlaGlyGluThrPro 424

```
QY 569 GCTTTTGCAGCATTGTTGATCTTAAC----- 595
Db 425 PheArgAlaAlaIleIleGlyThrArgGluIleAlaValProValIleSerMetThrIle 444
QY 596 ---TTGACTACAGTATTGGCTCAGCAGCTCTTTTCTTCTAGATACAGAGGCGCTATT--- 649
Db 445 AlaIleAlaValTyrSerProMetAlaLeuMetGlyGlyIleThrGlyThrLeuPhe 464
QY 650 AAAGGGTTGCTTTCACATGATTGAGAAATTTCTCTCTCAATGTTTACGGCTCTTTTC 709
Db 465 LysGluPheAlaLeuThrLeuAlaGlyAlaValPheIleSerGlyValValAlaLeuThr 484
QY 710 ATGACTAAATTTTCTTCATG-----CTGTGGATG--- 739
Db 485 LeuSerProMetMetSerSerLysLeuLeuLysSerAsnAlaLysProThrTyrMetGlu 504
QY 740 AATAAGACCCCAACATACA-----CAGTTG 763
Db 505 GluArgValGluHisThrLeuGlyLysValAsnArgValTyrGluTyrMetLeuAspLeu 524
QY 764 CATATGATGAATAAGTTTCGTGGGATAAAGCATGATTTCTTGAGAGGATGCAAAAACTT 823
Db 525 ValMetLeuAsnArg-----LysSerMet 532
QY 824 TGGGCTGTTCTGGAAGTGTTTTCTTTTA----- 853
Db 533 LeuAlaPheAlaValIlePheSerThrLeuProPheLeuPheAsnSerLeuSerSer 552
QY 854 -----GGTTGCGTTCCTCGGCTTCGGGTTTGGAGCGCTGGNAT 886
Db 553 GluLeuThrProAsnGluAspLysGlyAlaPheIleAlaIleGly---AsnAlaProSer 571
QY 887 TCCGTT----- 892
Db 572 SerValAsnValAspTyrIleGlnAsnAlaMetGlnProTyrMetLysAsnValMetGlu 591
QY 893 -----TTGGGAATGATTTTAAAGAGGAGGTATGCTTTACCTTTAAATCCA 937
Db 592 ThrProGluValSerPheGlyMetSerIleAlaGly---AlaProThrSerAsnSer 609
QY 938 AAGAGCATGCCATCAGC-----GATGTTGCTCAATGCGTGGCAAGTT 982
Db 610 SerLeuAsnIleIleThrLeuLysAspTyrLysGluArgSerArgLysGlnSerAlaIle 629
QY 983 GTGCATAAATACAGGAGCTGGTCTTCTCTAGAGACTTCCGTAATCAACATTT--- 1039
Db 630 MetAsnGluIleAsnGluLysAlaLysSerIleProGluValSerValSerAlaPheAsn 649
QY 1039 ----- 1039
Db 650 IleProGluIleAspThrGlyGluGlnGlyProProValSerIleValLeuLysThrAla 669
QY 1040 -----GGATCTTCAGAAAAGATC----- 1057
Db 670 GlnAspTyrLysSerLeuAlaAsnThrAlaGluLysPheLeuSerAlaMetLysAlaSer 689
QY 1058 ---AAAATCTATTTAGTGAATAAGCTTTAAGCTAT---ACTAAGCAGATACGAGCGCTCT 1111
Db 690 GlyLysPheIleTyrThrAsnLeuAspLeuThrTyrAspThrAlaGlnMetThrIleSer 709
QY 1112 CTCCTAAAA-----TTAAGCATCATG----- 1132
Db 710 ValAspLysGluLysAlaGlyThrTyrGlyIleThrMetGlnIleSerAsnThrLeu 729
QY 1133 -----AGCTGGCGT 1141
Db 730 GlySerPheLeuSerGlyAlaThrValThrArgValAspValAspGlyArgAlaTyrLys 749
QY 1142 TATTGTGGATTGTGTGAGAAAC-----AGCCCTAGATTT 1177
Db 750 ValIleSerGlnValLysArgAspArgLeuSerProGluSerPheGlnAsnTyrTyr 769
```

RESULT 12

```
US-09-912-020-282
; Sequence 282, Application US/09912020
; Patent No. US20020045592A1
; GENERAL INFORMATION:
; APPLICANT: Zyskind, Judith
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Trawick, John
; APPLICANT: Forsyth, R. Allyn
; APPLICANT: Froelich, Jamie M.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: GENES IDENTIFIED AS REQUIRED FOR PROLIFERATION IN
; FILE REFERENCE: ELITRA.001DV1
; CURRENT APPLICATION NUMBER: US/09/912,020
```

```

; CURRENT FILING DATE: 2001-07-23
; PRIOR APPLICATION NUMBER: 09/492,709
; PRIOR FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: 60/117,405
; PRIOR FILING DATE: 1999-01-27
; NUMBER OF SEQ ID NOS: 485
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 282
; LENGTH: 1040
; TYPE: PRT
; ORGANISM: E. coli
US-09-912-020-282

Alignment Scores:
Pred. No.: 8.62e-05 Length: 1040
Score: 129.00 Matches: 102
Percent Similarity: 38.45% Conservative: 81
Best Local Similarity: 21.43% Mismatches: 180
Query Match: 3.69% Indels: 114
DB: 10 Gaps: 18

US-09-868-987-1 (1-1864) x US-09-912-020-282 (1-1040)
QY 308 TTGGCAGATGCTTATTTGATGAGCGATATTAATTAATTT-----GAGGCC 355
DB 349 MetAlaIleAlaIleValMetIleIleIleuPheLeuArgAsnIleProAlaThr 368
QY 356 GTCAATCGCTTCGGAGAGCTGTTCTTCTGAAATCTTTGCTTATCTGAGCAGCTCTACAGTAT 415
DB 369 IleIleProGlyValAlaValProLeuSerIleGlyThrPheAlaValMetValPhe 388
QY 416 TTGATGCGGCACCTACCTTGTCAAGACTCGCTGGAGATGTTCTTGCATAGGCGATGCC 475
DB 389 LeuAspPheSerIleAsnAsnLeuThrIleMetAlaLeuThrIleAlaThrGlyPheVal 408
QY 476 GTATAGCAATGTTCTTGTATTCGAAAGATC-----CGAGAGCAATTT 520
DB 409 ValAspAspAlaIleValAlaIleGluAsnIleSerAlaGlyIleGluGlyGlyLys 428
QY 521 TTATGTCCTCAAGCTTTAAATAATCTGTAGAAAAGGATATACCAAGCTTTTGGAGCC 580
DB 429 ProLeuAlaIleAlaIleLeuIleGlyAlaGlyIleGlyPheThrIleIle----- 445
QY 581 ATTTTGATTTAACTTACTACAGATATGGCTTCAGACATTTCTTTTCTTCTGATACA 640
DB 446 -----SerLeuThrPheSerLeuIleAlaValIleProLeuLeuPheMet 461
QY 641 GGGCCTATTAAAGG-----TTTGCTTGACATTTGATTTGAAATTTTC 685
DB 462 GlyAspIleValGlyArgLeuPheArgGluPheAlaIleThrLeuAlaValAlaIleLeu 481
QY 686 TCTTCATGTTTACGGCTCTTTTCAGTAAATTTTCTTCATGCTGTGAGTAAGTAAAG 745
DB 482 IleSerAlaValAlaSerLeuThrLeuThrPro-----MetMetCysAlaArgMet 498
QY 746 ACCCAACATACACAGTTTCATATGATGATTAAGTTCGGGGGATTAACATGATTTCTTG 805
DB 499 LeuSerGlnGluSerLeuArgLysGlnAsnArgPheSerArgAlaSerGluLysMetPhe 518
QY 806 -----AGAGAGTCAAAAACCTT-----TGGGCT 829
DB 519 AspArgIleIleAlaAlaIleTyrGlyArgGlyLeuAlaLysValIleAsnHisProThrLe 538
QY 830 GTTTCGGAAGTGTCTTTCTTTAGGTGCGTTGCTCTCGGGATTTGAGCCTGGAATTC 889
DB 538 uThrLeuSerAla-----AlaLeuSerThrLeuLeuLeuSerValLe 552
QY 890 GTTTTGGGAATGATTTTAAAGAGGGTATGCTTTTACT-----TTAATTCAAA 940
DB 552 uLeuThrValPheIleProLysGlyPhePheProValGlnAspAsnGlyIleIleGlnG 572
QY 941 GAGCATGCAATCAGGAGATGTTGCTCAATGCGTGGCAAAAGTTTGCAATTAACATACAGAA 1000
; ; ; ; ;

```

```

DB 572 YThrLeuGlnAlaPro----- 577
QY 1001 GCTGCTCTTTCTTCTAGACCTTCGGATTCAAAACATTTGGATCTTCAGAAAAGATCAA 1060
DB 578 -----GlnSerSerPheAlaAsnMetAlaGlnArgGlnArgLysVal 592
QY 1061 ATCATTTTAGTATAAGCTTAAAGCTTAACTTAACATTAACACAGATAGAGCCCTCTCCATAAA 1120
DB 592 AlaAspValIleLeuGlnAspProAlaValGlnSer----- 604
QY 1121 TTAACGATCATGAGCTGCGTTATGTTGGATTTGTTTCAGAAACAGCCTAGATTTCTC 1180
DB 605 -----LeuThrSerPheValGlyValAspGlyThr 614
QY 1181 TACGAAAATCTTAAGCA-----AACGAAAATTTTGTCAAAAGTTAG 1224
DB 614 rAsnProSerLeuAsnSerAlaArgLeuGlnIleAsnLeuLysProLeuAspLysVal 634
QY 1225 CAGCAAACTTCGAAGAAATG---CGTTATCAGGCGACATCGGCTTTAGAGCTTT 1281
DB 634 pAspArgValGlnLysValIleAlaArgLeuGlnThrAlaValAspLysValProGlyVal 654
QY 1282 GGCATATCATCTTG-----CTCTATGTGAGTTTGGCGCTT 1314
DB 654 IaAspLeuPheLeuGlnProThrGlnAspLeuThrIleAspThrGlnValSer---ArgThr 673
QY 1315 TGAATGGCAATATGCTTTCAGTCCCGTATGCGCTTAATTCATGACCTTTTGGCTACCTG 1374
DB 673 rGlnTyrGlnPheThrLeuGlnAlaThrSerLeu-----AspAlaLeuSerThr-- 689
QY 1375 TGCAGCTCTGTTTATAGCAATTTCTTTTGAAGAAAATCAATATGATTTGCAAGCGAT 1434
DB 690 -----TrpValProGlnLeuMetGluLysLeuGlnGlnLe 701
QY 1435 TGGTGCTTAAATGATGCTATTTGGGATTCATTAACATATCTTGATCTTTGATCG 1494
DB 701 uProGlnLeuSerAspValSerSerAspTrpGlnAspLysGlyLeuValAlaTyrValAs 721
QY 1495 TATTCGTGAAGATCGCCAGACGAACTGTTTACCCTATGCAATGTTTAACTAATGATGC 1554
DB 721 nValAspArgAspSerAlaSerArgLeuGlyIleSerMet---AlaAspValAspAsnAl 740
QY 1555 CTTTCAAAAAGCTTCAGCGCGCAAGGTAATGACACAGCTACACT 1600
DB 740 aLeuTyrAsnAlaPheGlyGlnArgLeuIleSerThrIleTyrThr 755

RESULT 13
US-09-815-242-10203
; Sequence 10203, Application US/09815242
; Patent No. US2002061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselebeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625

```



```

; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 1410
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11516
; LENGTH: 1028
; TYPE: PRT
; ORGANISM: Helicobacter pylori
US-09-815-242-11516

Alignment Scores:
Pred. No.: 0.00111 Length: 1028
Score: 128.00 Matches: 73
Percent Similarity: 43.10% Conservative: 77
Best Local Similarity: 20.58% Mismatches: 142
Query Match: 3.86% Indels: 56
DB: 10 Gaps: 17

US-09-868-987-1 (1-1864) x US-09-815-242-11516 (1-1028)
QY 131 TTGAAATATCATGCGATGCTCGAGGAATTTACCCAGCGTGAAGTAGCAACTCGCC 190
DB 296 LeuYsh1st1eglnAl1leSerProserTyrlu1leatpPropheleuAspThr 315
QY 191 TCAGATTTAAATCTGAGCGATGCTTTTGTCCGAGTTCTCAGTGAAGACGATC 250
DB 316 SerTyrl1ealgnr-----Ser1legluAspVal 325
QY 251 TCTTCTGATCTTGGGAAAACAATGTCACAGAGCATATTCAGCATGCTGTGGCTG 310
DB 326 LysPheAspLeu-----lleuug1yAla1le-----leu 335
QY 311 GCATGCTTATTTGATGATGAGCGTATATTATGATTGGAGCGTC-----ATCGCT 364
DB 336 Ala1leVal1leValPhe-----AlaPheleuAspSerGlyThr1leThrLeuValSer 353
QY 365 TCGGAGCGTGTCTGATCTGATCTTGTGCTTATCTGGGAGCTCTACAGATTTGGATCG 424
DB 354 Ala1leSer1lePro1leSer1leMetGlyThrPheAla1leu1leGlnTrpMetGlyPhe 373
QY 425 CCATTCACCTTGTGACGCTGCGGAGTTGTTCTGTATGGGATGCGCGTAGATGCA 484
DB 374 SerLeuAsnMetLeuThrMetValAla1leuThrLeuAla1leGly1le1leAsp 393
QY 485 AATGTTCTTATTCGAGAGATCCGAGAAATTTTATGTCCAAGCTTAAAAA 544
DB 394 Ala1leVal1leGlnAsn1leHis1ySylsLeuGluMetGlyMetSer1yAspGly 413
QY 545 TCTGTAGAAAAGATATACCAAGGCTTTGGAGCCATTTTGATTTCACTTGACTACA 604
DB 414 AlaSerTyrgluGly---Val1ySglu1leGlyPheAla-----LeuValAla 428
QY 605 GATTTGCGCTGACACTTCTTTCTCTACATACAGGCGCTATTAAAGG----- 655
DB 429 IleSerAlaMetLeuLeuSerValPheValPro1leGlyAsnMet1ySyl1le1leGly 448
QY 656 -----TTTGCTTGAACATTTAGGATTTTGAAGATTTCTCTCAATGTTTACG 700
DB 449 ArgPhePheGlnSerPheGly1leThrValAla1leuAla1leuSerTyrglyVal 468
QY 701 GCTCTTTCATGACTAAATTTTCTTCATGCTGTGAGATGATTAAGCCACATACACAG 760
DB 469 ValVal1Thr1le1lePheMetValSerSerValValAsn---ProArg1SerArg 487
QY 761 TTGCATATGATGAATAC-----TTGCGGGGATTAACATGATTTCTTGAGAGATCG 814
DB 488 PheTyMetTrpSerGluProPhePhe1yAla1leuGlnSerArgTyrlThr1ySleu 507
QY 815 AAAAACTTTGGGCTGTTCTCGAAGTGTTTTCTTTAGTTGCTGCTGCTCGGGTTT 874
DB 815 AAAAACTTTGGGCTGTTCTCGAAGTGTTTTCTTTAGTTGCTGCTGCTCGGGTTT 874

```

```

DB 508 Gln-----TrpVal1leuAsn1ySylsLeu1le1lePhe1leAlaVal1leuValPhe 525
QY 875 -----GGAGCGCTG-----AATCCGTTTGGGATGATTTT-----AAA 910
DB 526 ValGlySerLeuPheValAlaSer1ySylsLeuGlyMetGluPheMetLeu1ySgluAspArg 545
QY 911 GAGGATATGCTTTTACCTTTAATCCAAAAGACATGGCATGACG---GATGTTGCTCA 967
DB 546 GlyArgPheLeuValTrpLeu1ySylsPro---GlyValSer1leAspTyrlMetThr 564
QY 968 ATGCGTGGCAAGTT-----GTGCATTAACCTACAGAACTGCTTCTTCTTCT 1015
DB 565 Gln1ySer1ySer1lePheGln1ySylsAla1leGlu1ySylsAspGluValGluPheThr 584
QY 1016 AGAGACTTCGCTTCAACATTTGATGCTTCGAAAAGATCAAAATCTATTATGAT 1075
DB 585 LeuGlnVal1yGlyTyrlGlyThrGlnAsnProphe1yAla1y1lePheValGlnLeu 604
QY 1076 AAGCTTTAAGCTATTAAGCAG 1099
DB 605 LysProLeu1ySgluArg1ySyls 612

RESULT 15
US-09-912-020-283
; Sequence 283, Application US/09912020
; Patent No. US20020045592A1
; GENERAL INFORMATION:
; APPLICANT: Zyskind, Judith
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Trawick, John
; APPLICANT: Forsyth, R. Allyn
; APPLICANT: Froelich, Jamie M.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: GENES IDENTIFIED AS REQUIRED FOR PROLIFERATION IN
; FILE REFERENCE: ELITRA.001DVI
; CURRENT APPLICATION NUMBER: US/09/912,020
; CURRENT FILING DATE: 2001-07-23
; PRIOR APPLICATION NUMBER: 09/492,709
; PRIOR FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: 60/117,405
; PRIOR FILING DATE: 1999-01-27
; NUMBER OF SEQ ID NOS: 485
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 283
; LENGTH: 1025
; TYPE: PRT
; ORGANISM: E. Coli
US-09-912-020-283

Alignment Scores:
Pred. No.: 0.000318 Length: 1025
Score: 123.50 Matches: 130
Percent Similarity: 33.33% Conservative: 105
Best Local Similarity: 18.44% Mismatches: 243
Query Match: 3.73% Indels: 227
DB: 10 Gaps: 28

US-09-868-987-1 (1-1864) x US-09-912-020-283 (1-1025)
QY 245 ACGATCTCTCTGATCTTGGGAAAACAATGTCACAGAGCATTTATCTGACATGCTGT 304
DB 326 Thr1leArgAlaSerLeuGluGluValGluGlnThrLeu---lle1leSer----- 341
QY 305 GCGTGGCAATGCTTATGTTTGTGATGAGCGTATATTATGATTTGA-----GGCGTC 358
DB 342 ---ValAlaLeuVal1leLeuValValPheLeuPheLeuAspSerGlyArgAlaThr1le 360
QY 359 ATCGCTTGGGAGCTGTTCTTCTGATCTTGTGCTTATCTGAGAGCTCTACATATTG 418
DB 361 lleProAlaValSerValProValSer1le1yThrPheAla1leMetTyrlLeuGly 380

```


QY 419 GATCGCCCACTACCTTGTCTCAGGACTCGCTGGGATTGTTCTTGTCTATATGGGATGGCCGTA 478
Db GlyPheSerLeuAsnAsnLeuSerLeuMetAlaLeuThrIleAlaThrGlyPheValVal 400
QY 479 GATGCAAAATGTTCTTATTCGAAAGAAATCCGAGAGGAATT-----TTA 523
Db AspAspAlaIleValValLeuGluAsnIleAlaArgHisLeuGluAlaGlyMetLysPro 420
QY 524 TTGCTCAAGCTCTTAAATAATCTGTAGAAAAGGATATACCAAGGCTTTTGGAGCCATT 583
Db LeuGlnAlaLeuGlnGlyThrArgGluValGlyPheThr-----Val 435
QY 584 TTGATTCTAATCTACGATACAGTATGGCCCTCAGCAGCTCTTTTCTTCTAGATACAGGG 643
Db LeuSerMetSerLeuSerLeuValAlaValPheLeuProLeuLeuMetGlyGlyLeu 455
QY 644 CCT-----ATTAAAGGGTTCTTGTGATGATTAGGAATTTCTCTTCAATG 694
Db ProGlyArgLeuLeuArgGluPheAlaValThrLeuSerValAlaIleGlyIleSerLeu 475
QY 695 TTACGGCTCTTTTCATGACTAAATTTTCTTCTCATGCTGTGATGAATAAGACC----- 748
Db LeuValSerLeuThrLeuThrPro---MetMetCysGlyTrpMetLeuLysAlaSerLys 494
QY 749 -----CAACATACACAGTTGCATATGATGAATAAG-----TTCGTGGGATGAAGCATGAT 799
Db ProArgGluGlnLysArgLeuArgGlyPheGlyArgMetLeuValAlaLeuGlnGlnGly 514
QY 800 TTCCTGAGGAGTGCACAAAATCTTGGGCTGTTCTCGAAGT-----GTTTTT 847
Db TyrGlyLysSerLeuLys-----TrpValLeuAsnHisThrArgLeuValGlyValVal 532
QY 848 CTTTGTAGTGTGGTCTCTCGGTTTGGAGCTCGAATTCGGTT----- 892
Db LeuLeuGlyThrIleAlaLeuAsnIleTrpLeuTyrlleSerIleProLysThrPhePhe 552
QY 893 -----TTGGGA----- 898
Db ProGluGlnAspThrGlyValLeuMetGlyGlyIleGlnAlaAspGlnSerIleSerPhe 572
QY 899 -----ATGGATTT----- 907
Db GlnAlaMetArgGlyLysLeuGlnAspPheMetLysIleIleArgAspProAlaVal 592
QY 908 -----AAAGGAGGTATGCCCTTTACC 928
Db AspAsnValThrGlyPheThrGlySerArgValAsnSerGlyMetMetPheIleThr 612
QY 929 TTTAATCAAAGACATGGCATCAGCGATGTGTCAAAATCGGTGGCAAAAGTTGTGCAT 988
Db LeuLysProArgAspGluArgSerGluThrAlaGlnGlnIleIleAspArgLeuArgVal 632
QY 989 AAACATA---CAGGAGCTGTT-----CTTCTCTAGACCTTCGCTATT 1030
Db LysLeuAlaLysGluProGlyAlaAsnLeuPheLeuMetAlaValGlnAspIleArgVal 652
QY 1031 -----CAACATTTGGATCTTCACAAAGATCAAAATCTATTTTAGTGATAAGCTTTA 1084
Db GlyGlyArgGlnSerAsnAlaSerTyrlleThrLeuLeuSerAspAspLeuAlaAla 672
QY 1085 -----AGCTATACAGCATACGACCTCTCTCTAAATTAACGATCATG----- 1132
Db LeuArgGluTrpGluProLysIleArgLysLysLeuAlaThrLeuProGluLeuAlaAsp 692
QY 1133 -----AGC 1135
Db ValAsnSerAspGlnAspAsnGlyAlaGluMetAsnLeuValTyrlleAspArgAspThr 712
QY 1136 TGGCGTTATTGGGATTTGTTCTCAGAACAGCGCTAGATTCTCTACGGAACCTCTAAA 1195
Db MetAlaArgLeuGlyIleAspValGln-----AlaAlaAsnSerLeu 726

QY 1196 CGAACCCAAATTTTGGTCAAGGTAAGCAGCAAACTATCGAAGAAATG----- 1246
Db LeuAsnAsnAlaPheGlyGlnArgGlnIleSerThrIleTyrlleProMetAsnGlnTyrlle 746
QY 1247 -----CGTTATCAGCGCACCATCGGCTTTTAGGAGCTTTG 1282
Db LysValValMetGluValAspProArgTyrlleThrGlnAspIleSerAlaLeuGluLysMet 766
QY 1283 GCAATCATC-----TTGCTCTATGTAGTTTGGCCTTTTGAATGCAATAT 1327
Db PheValIleAsnAsnGluGlyLysAlaIleProLeuSerTyrllePheAlaLysTrpGlnPro 786
QY 1328 GCT----- 1330
Db AlaAsnAlaProLeuSerValAsnHisGlnGlyLeuSerAlaAlaSerThrIleSerPhe 806
QY 1330 ----- 1330
Db AsnLeuProThrGlyLysSerLeuSerAspAlaAlaIleAspArgAlaMetThr 826
QY 1331 -----TTCAGTCCGCTATCGCTTTTAAATTCAT 1357
Db GlnLeuGlyValProSerThrValArgGlySerPheAlaGlyThrAlaGlnValPheGln 846
QY 1358 GACCTTTTGGCTACTCTGCAGTCTTGTATATACACATCTTCTTTTGAAGAAATTCAA 1417
Db GluThrMetAsnSerGlnValIleLeuIleAlaIleAlaIleAlaThrValTyrlleVal 866
QY 1418 ATAGATTG----- 1426
Db LeuGlyIleLeuTyrlleGluSerTyrlleValHisProLeuThrIleLeuSerThrLeuProSer 886
QY 1427 CAAGCCATTGTGCTTTA----- 1444
Db AlaGlyValGlyAlaLeuLeuAlaLeuLeuLeuPheAsnAlaProPheSerLeuIleAla 906
QY 1445 -----ATGACTGTATTCGGGTATTCATAACAATATCTTTGATCATTTTGTGAT 1492
Db LeuIleGlyIleMetLeuLeuIleGlyIleValLysLysAsnAlaIleMetMetValAsp 926
QY 1493 CGTATTCGTGAAGATCGCAAGCGAACCTGTTTACCCCTATGCATGTTTATAGTTAATGAT 1552
Db PheAlaLeuGluAlaGlnArgHisGlyAsnLeuThrProGlnGluAlaIlePheGlnAla 946
QY 1553 GCCCTTCAAAGACGTTTCAGCGCACGTAATGACACAGCTACAACTCTATCAGTTTGTG 1612
Db CysLeuLeuArg---PheArgProIleMetMetThrThrLeuAlaAlaLeuPheGlyAla 965
QY 1613 TTAATGCTTTTGTATAGCGGC---TCCTCTCTCTTTTAAATTTTGTGATTTATATGACC 1669
Db LeuProLeuValLeuSerGlyAspGlySerGluLeuArgGlnProLeuGlyIleThr 985
QY 1670 ATA-----GGATTCCTTAGGAACCTTTATCGTCTCTTTATATGACCACTCTGTTG 1723
Db IleValGlyGlyLeuValMetSerGlnLeuLeuThrLeuTyrlleThrProValValTyrlle 1005
QY 1724 TTGTTTATGTCCTG 1738
Db LeuPhePheAspArg 1010
RESULT 16
US-09-815-242-5111
; Sequence 5111, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard

```

; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5111
; LENGTH: 1062
; TYPE: PRF
; ORGANISM: Pseudomonas aeruginosa
US-09-815-242-5111

```

Alignment Scores:

```

Pred. No.: 0.00408 Length: 1062
Score: 122.50 Matches: 65
Percent Similarity: 39.86% Conservative: 51
Best Local Similarity: 22.34% Mismatches: 114
Query Match: 3.70% Indels: 61
Gaps: 11

```

US-09-868-987-1 (1-1864) x US-09-815-242-5111 (1-1062)

```

QY 104 GTCAGACCCCTATTATTAACGTCCTCAATGCAAAATCATGCCAGTGTCTCAGGAAATTT 163
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 290 ValAlaIleProIlePheGlnArgProGlySerAsnAlaIleGluIleSer---AsnLeu 308
QY 164 ACCACCGTGAAGAGCAAACTGCCCAATTTAAATGAGAGGATGCTTTGTT 223
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 309 ValArgGluSerMetAlaGluLeuHisSerPheProGlnGly---MetAspTyr--- 326
QY 224 CCCAGGTTCTCAGTGAAGAGCAGCATCTTGTATCTTGGGAAAAACAATGTACACA 283
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 327 ---SerIleValTyrAspProThiIle-----PheValArg 337
QY 284 GGCATTTATCTCAGCA-----TGCTGTGCTTGGCATGCTTATT 322
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 338 GlySerIleGluAlaValAlaHisThrLeuPheGluAlaLeuValIleValIleVal 357
QY 323 GTTTGATGACCGTATATTAAGATTGGAGCGCATCGCTTGGGAGCTGTTCTTCTG 382
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 358 ValIleLeuPheLeuGlnThrTyrPArgLaserIleIleProLeuAlaIleValProVal 377
QY 383 AATCTTTGCTTATCTATCTGAGCCTCTACAGATTGAGTGGCAGCATCACTTGTCAGA 442
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 378 SerIleuIleGlyThrPheAlaValMetHisMetLeuGlyPheSerLeuAsnAlaLeuSer 397
QY 443 CTCGCTGGATGTTGTTCTGATAGGGAGTGGCCGTAGATGCAAAATGTTCTTGATTCGA 502
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 398 LeuPheGlyLeuValIleValIleAlaIleGlyIleValIValAspAspAlaIleValIleValGlu 417
QY 503 AGAATCCGAGAGGAATTTTATGTGCTCAAAAGTCTTAAATAATCTGTGAATAAAGAGAT 562
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 418 AsnValGluArgAsnIleGlyLeuGlyLeuLysProValGluAlaThrLysArgAlaMet 437
QY 563 ACCAAGGCTTTGGAGCATTTTGAATCTTAACCTTGAAGTATGATGAGTGGCTTGGCACTT 622
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 438 ArgGluValThrGlyProIle-----IleAlaIThrAlaLeu 449

```

```

QY 623 CTTTTC-----TTCCAGATACAGGCGCTATTAAAGG----- 655
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 450 ValLeuCysAlaValPheIleProThrAlaPheIleSerGlyLeuThrGlyGlnPheTyr 469
QY 656 -----TTTGCTTTGACATGATTTTGAAGATTTTCTTTCATATGTTTACGGCTTTTC 709
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 470 ArgGlnPheAlaLeuThrIleAlaIleSerThrValIleSerAlaPheAsnSerLeuThr 489
QY 710 ATGACT-----AAATTTTCTTC 727
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 490 LeuSerProAlaLeuAlaValLeuLeuLysGlyHisIleGluProLysAspArgPhe 509
QY 728 ATGCTGGATGATTAAGACCAACATACAGTTG---CATATGATGATTAAGTTC--- 781
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 510 SerValPheLeuAspLysLeuLeuGlySerTyrPheArgPheArgPheAsnArgPhePhe 529
QY 782 -----GTGGGATTAAGCATGATTTTCTTGAAGATGCAAAAA 820
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 530 AspArgAlaSerHisGlyTyrValGlyThrValAsnArgValLeuArgLysSerSerIle 549
QY 821 CTTGGGCTGTTTCTGGAAGTGTCTTTTCTTTA 853
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 550 AlaLeuLeuValTyrGlyGlyLeuMetValLeu 560

```

RESULT 17

US-09-895-913A-10
Sequence 10, Application US/09895913A
Patent No. US20020160456A1

GENERAL INFORMATION:

APPLICANT: Kleantous, Harold
APPLICANT: Al-Garawi, Amal
APPLICANT: Miller, Charles
APPLICANT: Tomb, Jean Francois
APPLICANT: Oomen, Raymond P.
TITLE OF INVENTION: Identification of Polynucleotides
TITLE OF INVENTION: Encoding No. US20020160456A1 Helicobacter Polypeptides in th
FILE REFERENCE: 06132/043002
CURRENT APPLICATION NUMBER: US/09/895,913A
CURRENT FILING DATE: 2001-06-29
PRIOR APPLICATION NUMBER: US 08/881,227
PRIOR FILING DATE: 1997-06-24
NUMBER OF SEQ ID NOS: 368
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 10
LENGTH: 761
TYPE: PRF
ORGANISM: Helicobacter pylori
US-09-895-913A-10

Alignment Scores:

```

Pred. No.: 0.000399 Length: 761
Score: 122.00 Matches: 57
Percent Similarity: 45.52% Conservative: 75
Best Local Similarity: 19.66% Mismatches: 124
Query Match: 3.68% Indels: 34
Gaps: 13

```

US-09-868-987-1 (1-1864) x US-09-895-913A-10 (1-761)

```

QY 305 GGCTTGCCATGCTTATTTGATGAGCGTATATTATGATTTGGAGCGTC----- 358
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 332 GlyAlaIleLeuAlaValLeuValAlaPheAlaPheLeuArgAsnGlyThrIleThrLeu 351
QY 359 ATGCTTCGGAGCTGTTCTTGAATCTTTGCTTATCTGAGCAGCTCTACAGTATTTG 418
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 352 ValSerAlaIleSerIleProIleSerIleMetGlyThrPheAlaLeuIleGlnTrpMet 371
QY 419 GATGCGCACTACCTTGTGAGACTGCTGAGATTGTTCTTGTATGGGATGCGCGTA 478
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 372 GlyPheSerLeuAsnMetLeuThrMetValAlaLeuThrLeuAlaIleGlyIleIleIle 391
QY 479 GATGCAATGTCTTGTATTCGAAAGATCCGAGAGCAATTTTATTTGTCTCAAGTCTT 538

```

```

Db 392 AspaspAlaIleValIleGluAsnIleHisLysLysLeuGluMetGlyMetSerLys 411
Qy 539 AAAAAATCTAGAAAAAGGATATACCAAGGCTTTTGGAGCCATTTTGTATCTTAACCTTG 598
Db 412 ArgLysAlaSerTyrGluGlyValArgGluIleGlyPheAla-----Leu 426
Qy 599 ACTACAGATTGGCTCAGCACTCTTTCTCTCTAGATACAGGCCCTATTAAAGGG--- 655
Db 427 ValAlaIleSerAlaMetLeuLeuSerValPheValProIleGlyAsnMetLysGlyIle 446
Qy 656 -----TTTGCTTTGACATTGATTTTAGGAATTTCTCTTCATG 694
Db 447 IleGlyArgPhePheGlnSerPheGlyIleThrValAlaLeuAlaIleAlaLeuSerTyr 466
Qy 695 TTTACGGCTCTTTTCATGACTAAATTTTCTTCATGCTGTGTGATGAATPAAGACCAACAT 754
Db 467 ValValValValThrIleIleProMetValSerSerValValValAsn---ProArgHis 485
Qy 755 ACACAGTTGCATATCATGAATAAG-----TTCGTGGGGATAAAGCATGATTTCTTGAGA 808
Db 486 SerArgPheTyrValTrpSerGluProPhePheLysAlaLeuGluSerArgTyrThrLys 505
Qy 809 GGATCAAAAAAATTTTGGCTGTCTCTCGAAGTGTCTTTTATAGGTGGCTGCTCTC 868
Db 506 LeuLeuGln-----TrpValLeuAsnHisLysIleIleIleSerIleAlaValLeu 523
Qy 869 GGGTTT---GGAGCTCG---AATTCCGTTTTTGGGAATGATTTT----- 907
Db 524 ValPheValGlySerLeuPheValAlaSerLysIleGlyMetGluPheMetLeuLysGlu 543
Qy 908 ---AAAGAGGTATCGCTTTACCTTTTAAATCCAAAGAGCATGGCATCAGC---GATGTT 961
Db 544 AspArgGlyArgPheLeuValTrpLeuLysAlaLysPro---GlyValSerIleAspTyr 562
Qy 962 GCTCAATGCTGGCGCAAGTGTGTCATAAATCTACAGGAAGCTGGTCTTCTCTAGAGAC 1021
Db 563 MetThrGlnLysSerLysIlePheGlnLysAlaIleGluLysHisAlaGluValGluPhe 582
Qy 1022 TTCCGTATTCAACA---TTTGGATCTTTCAGAA-----AAGATCAAAATCTATTATT 1069
Db 583 ThrThrLeuGlnValGlyTyrGlyThrGlnAsnProPheLysAlaLysIlePheVal 602
Qy 1070 AGTGATAAGCTTTAAGCTATCTACTAAGCAG 1099
Db 603 GlnLeuLysProLeuLysGluArgLysLys 612

```

RESULT 18

```

US-09-815-242-5194
; Sequence 5194, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIORITY FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIORITY FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIORITY FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIORITY FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIORITY FILING DATE: 2000-10-23

```

```

; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5194
; LENGTH: 1043
; TYPE: PRF
; ORGANISM: Pseudomonas aeruginosa
US-09-815-242-5194

Alignment Scores:
Pred. No.: 0.000456 Length: 1043
Score: 122.00 Matches: 154
Percent Similarity: 33.83% Conservative: 117
Best Local Similarity: 19.23% Mismatches: 260
Query Match: 3.68% Indels: 270
DB: 10 Gaps: 37

US-09-868-987-1 (1-1864) x US-09-815-242-5194 (1-1043)
Qy 74 CGTATGCTGTAGTGTGACGGTTATATGTCAGCAGCCCTATTTTAAAGCTCCCATG 133
Db 269 ArgLeuGluLeuGlyLysGluSerTyrAsnIleSerArgLeuAsnGlyThrPro--- 287
Qy 134 AAAAATCATGCGATGCTCAGGGGAAATTTACCCACCGTGAAGTGAGCAAACTCGCTCA 193
Db 288 -----ThrValGlyAlaIle-----GlnLeuSerProGlyAlaAsn 300
Qy 194 GATTAAATCTGGAGCGATG-----TCTTTTGTTCCTC 226
Db 301 AlaIleGlnThrAlaThrLeuValLysGlnArgLeuAlaGluLeuSerAlaPhePro 320
Qy 227 GAG-----GTTCTCAGTGAAGAGCAGATCTCTCTGATCTCTGGGAAAAA 271
Db 321 GluAspMetGlnTyrSerValProTyrAspThrSerArgPheValAspValAlaIleGlu 340
Qy 272 CAATGTACACAAAGCATTTCTCAGCA-----TGCTGTGCTTGGCAATGCTTATTGTT 325
Db 341 LysValIleHisThrLeuIleGluAlaMetValLeuValPheLeuValMetPheLeuPhe 360
Qy 326 TTGATGACGCTATATTATAGATTTTGGAGGCGTCATCGCTTCGGGAGCTGTTCTTCTGAAT 385
Db 361 LeuGlnAsnValArgTyrThrLeuIleProSerIleValValProValCysLeuLeuGly 380
Qy 386 CTTTGTCTATCTGGGCGCTCTACAGTATTGATGGCCACTCACCCTTGTCTAGGACTC 445
Db 381 ThrLeuMetVal-----MetTyrLeuLeuGlyPheSerValAsnMetMetThrMet 397
Qy 446 GCTGGGATTGTTCTGTATGGGATGGCGGTAGATGCAAAATGTTCTTGTATTGCAAGA 505
Db 398 PheGlyMetValLeuAlaIleGlyIleLeuValAspAlaIleValValValGluAsn 417
Qy 506 ATCCGAGAGGAATTTTATTCTCTCAAAAGCTTTAAA-----AAATCTGTAGAAAAAGGA 559
Db 418 Val---GluArgIleMetAlaGluGluGlyLysSerProAlaGluAlaThrValLysAla 436
Qy 560 TATACCAAGGCTTTTGGAGCCATTTTGTATCTTAACCTTACAGTATGTCGCTCAGCA 619
Db 437 MetLysGlnValSerGlyAlaIleVal-----GlyIleThrLeuValLeuSerAlaVal 454
Qy 620 CTT-----CTTTCTTCCTA-----GATACAGGGGCTATT---AAAGGGTTTGTCTTG 664
Db 455 PheLeuProLeuAlaPheMetAlaGlySerValGlyValIleTyrGlnPheSerVal 474
Qy 665 ACATTGATTTTAGGAATTTTCTCTCTCAATGTTTACGGCTCTTTTCATGACTTAAA----- 718
Db 475 SerLeuAlaValSerIleLeuPheSerGlyPheLeuAlaLeuThrPheThrProAlaLeu 494
Qy 719 -----TTTTTTC--- 724

```



```

; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FASTSEQ for Windows
; SEQ ID NO: 5330
; LENGTH: 801
; TYPE: PR1
; ORGANISM: Staphylococcus aureus
US-09-815-242-5330

```

Alignment Scores:	
Pred. No.:	0.00104
Score:	18.00
Percent Similarity:	48.30%
Best Local Similarity:	24.43%
Query Match:	3.56%
DB:	10
Length:	801
Matches:	43
Conservative:	42
Mismatches:	73
Indels:	18
Gaps:	7

US-09-868-987-1 (1-1864) x US-09-815-242-5330 (1-801)

[illegible]

RESULT 20

US-09-815-242-10072
; Sequence 10072, Application US/09815242

```

; PATENT NO. US2902006/136941
; GENERAL INFORMATION:
; APPLICANT: Haseelbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA-011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21

```

	;	PRIOR APPLICATION NUMBER:	60/191,078	
	;	PRIOR FILING DATE:	2000-03-21	
	;	PRIOR APPLICATION NUMBER:	60/206,848	
	;	PRIOR FILING DATE:	2000-05-23	
	;	PRIOR APPLICATION NUMBER:	60/207,727	
	;	PRIOR FILING DATE:	2000-05-26	
	;	PRIOR APPLICATION NUMBER:	60/242,578	
	;	PRIOR FILING DATE:	2000-10-23	
	;	PRIOR APPLICATION NUMBER:	60/253,625	
	;	PRIOR FILING DATE:	2000-11-27	
	;	PRIOR APPLICATION NUMBER:	60/257,931	
	;	PRIOR FILING DATE:	2000-12-22	
	;	PRIOR APPLICATION NUMBER:	60/269,308	
	;	PRIOR FILING DATE:	2001-02-16	
	;	NUMBER OF SEQ ID NOS:	14110	
	;	SOFTWARE:	FastSeq for Windows Version 4.0	
	;	SEQ ID NO	10072	
	;	LENGTH:	1049	
	;	TYPE:	PRT	
	;	ORGANISM:	Escherichia coli	
	US-	09-815-242-10072		
	Alignment Scores:			
	Pred. No.:	0.00131	Length:	1049
	Score:	117.50	Matches:	138
	Percent Similarity:	30.95%	Conservative:	109
	Best Local Similarity:	17.29%	Mismatches:	278
	Query Match:	3.55%	Indels:	273
	DB:	10	Gaps:	33
	US-	09-868-987-1 (1-1864) x US-	09-815-242-10072 (1-1049)	
Qy	44	AATGGCAAAATTCTTGCAAAACCGGTAGTGCCTGATGGCTGTAGTCAGTCAGCGTTATATG	103	
Db	282	AsnGlyGlnProAlaserGlyLeuGlyIleLysLeuAlaThrGlyAlaAsnAlaLeuAap	301	
Qy	104	GTCAGCAGCCCTATTTTAAACGTGCCATCCCATGAAAATCATGCCAGTCTCTCAGGGAATTT	163	
Db	302	ThrAlaAlaIle	306	
Qy	164	ACCACCGTCAAGTGAGCAGCAACTCGCCTCAGATTAAAAATCTGGAGCGATGCTTTTGT	223	
Db	307	--ArgAlaGluLeuAlaLysMetGluProPhePheProSerGlyLeuLysIleValTy	325	
Qy	224	CCC-----GAGGTCTCAGTGAAGACGACGATCTCTCTGATCTT	262	
Db	326	ProTyraSpThrThrPropheValIylleSeriHsiGluVal	341	
Qy	263	GGGMAAACAATGTACAAAGCATTAATCTCAGCATGCTGGCTTGGCAATGCTTATT	322	
Db	342	----LysThrLeuValGluAlaIleIleLeuValPhe-----LeuValMetTy	357	
Qy	323	GTTTTGTAGAGCTATATTATAGATTGGGGCGTCATCGCTTCGGGAGCTGTTCTCTCG	382	
Db	358	PheLeuGlnAsnPheArgAlaThrLeuilleProThrIleAlaValProValValLeuLeu	377	
Qy	383	AATCTTTTGTCTTATCTGGCGAGCTCTACAGTATTGTCATGGTCGCGCACTCACCTTGT	442	
Db	378	GlyThrPheAlaValLeuAlaIlaPheGlyPheSerIleAsnThrLeuThrMetPheGly	397	
Qy	443	CTCGCTGGGATGTTCTTGTCTATGGGGATGGCGGTAGATGCAAAATGTTCTGTATTTCGAA	502	
Db	398	Met-----ValLeuAlaIleGlyLeuLeuValAspAlaIleValValValGlu	414	
Qy	503	AGAATCCGACGAGGAATTTTATTGTTCTCAAAGTCTT-----AAAAATCTGTAGAAAAA	556	
Db	415	AsnVal---GluArgValMetAlaGluGluLeuProProllysGluAlaThrArGLys	433	
Qy	557	GGATATACCAAGGCTTTTGGAGCCATT-----TTTGATTCTTAACITTGACTACAGTATTG	610	
Db	434	SerMetGlyGlnIleGlnGlyAlaLeuValGlyIleAlaMetValLeuSerAlaValPhe	453	
Qy	611	GCCTCAGCACTTCTTTCTTCTTAGATACAGGCGCCTATT---AAAGGGTTTGTCTTGACA	657	

```

Db 454 ValProMetAlaPhePheGlyGlySerThrGlyAlaIleTyrArgGlnPheSerIleThr 473
Qy 668 TTGATTTCAGCAATTTCTCTTCATAGTTTACGGCTTTTCATGACT----- 715
Db 474 IleValSerAlaMetAlaLeuSerValLeuValAlaLeuIleLeuThrProAlaLeuGly 493
Qy 716 -----AAATTTTCTTCATG 730
Db 494 AlaThrMetLeuLysPheIleAlaLysGlyAspHisGlyGluGlyLysGlyPhePhe 513
Qy 731 CTGGGATGATTAAGACCAACATACACAGTTGCATATGAAATGATGGTGGG--- 787
Db 514 GlyTyrPheAsnArgMetPheGluLysSerThrHisIleTyrThrAspSerValGlyGly 533
Qy 787 ----- 787
Db 534 IleLeuArgSerThrGlyArgTyrLeuValLeuTyrLeuIleIleValIleValGlyMetAla 553
Qy 787 ----- 787
Db 554 TyrLeuPheValArgLeuProSerSerPheLeuProAspGluAspGlnGlyValPheMet 573
Qy 787 ----- 787
Db 574 ThrMetValGlnLeuProAlaGlyAlaThrGlnGluArgThrGlnLysValLeuAsnGlu 593
Qy 788 ATAAAGCATGATTTCTTG-----AGAGATGCAAAAACTTTGGGCTGTTCT 835
Db 594 ValThrIleTyrTyrLeuThrThrLysGluLysAsnValGluSerValPheAlaValAsn 613
Qy 836 GGAAGTGTTCCTT-----TTAGTTCGCTTCCTCGGCTTGGAGCC 880
Db 614 GlyPheGlyPheAlaGlyArgGlyGlnAsnThrGlyIleAlaPheValSerLeuLysAsp 633
Qy 881 TGG-----AATCCGTTTGGGATG----- 901
Db 634 TrpAlaAspArgProGlyGluGlnAsnLysValGluAlaIleThrMetArgAlaThrArg 653
Qy 902 -----GATTTTAAAGGAGGTATGCTTTACCTTTAAT----- 934
Db 654 AlaPheSerGlnIleLysAspAlaMetValPheAlaPheAsnLeuProAlaIleValGlu 673
Qy 934 ----- 934
Db 674 LeuGlyThrAlaThrGlyPheAspPheGluLeuIleAspGlnAlaGlyLeuGlyHisGlu 693
Qy 934 ----- 934
Db 694 LysLeuThrGlnAlaArgAsnGlnLeuLeuAlaGluAlaLysHisPheAspMetLeu 713
Qy 935 -----CCAAAGAGCATGCGATCAGCATGTCCTCAATGCGTGGCAAGTT---GTG 985
Db 714 ThrSerValArgProAsnGlyLeuGlnAspThrProGlnPheLysIleAspIleAspGln 733
Qy 986 CATAACTACAGGAAGCTGCTCTTCTCTAGAGACTTCCTGATTTCAACCTTTGGA--- 1042
Db 734 GluLysAlaGlnAlaLeuGlyValSerIleAsnAspIleAsn---ThrThrLeuGlyAla 752
Qy 1042 ----- 1042
Db 753 AlaTrpGlyLysTyrValAsnAspPheIleAspArgGlyArgValLysLysValTyr 772
Qy 1043 -----TCTTCAGAAAGATCAAAATCTATTATGATTAAGCTTAAAGCTATACATAAG 1096
Db 773 ValMetSerGlnAlaLysTyrArgMetLeuProAspAspIleGlyAspTrpTyrValArg 792
Qy 1097 CAGATACAGGCTCTCTCAAAATTAACATCATGAC-----TGGCGTTATTGTT 1147
Db 793 AlaAlaAspGlyGlnMetValProPheSerAlaPheSerSerArgTyrGluTyr--- 811
Qy 1148 GGGATTGTTTCAGAAACAGGCTTATGATT----- 1177

```

```

Db 812 -----GlySerProArgLeuGluArgTyrAsnGlyLeuProSerMetGlu 826
Qy 1178 -----CTTACGGAACCTTAAACGAACCA---AAATTTGGTCAAAG 1219
Db 827 IleLeuGlyGlnAlaAlaProGlyLysSerThrGlyGlnAlaMetGluLeuMetGluGln 846
Qy 1220 GTAAGCAGCAAACTATCGAAGAAATCGCTTACAGCGCAGCATGGGCGTT----- 1270
Db 847 LeuAlaSerLysLeuProThrGlyValGlyTyrAspTrpThr---GlyMetSerTyrGln 865
Qy 1271 -----TTAGAGCTTTGGCAATCATCTTCTCTAT 1300
Db 866 GluArgLeuSerGlyAsnGlnAlaProSerLeuTyrAlaIleSerLeuIleValAlaPhe 885
Qy 1301 GTC-----AGTTGCGCTTGAATGCAATATGCTTACGTGCGCTATGCGCTTAA 1351
Db 886 LeuGlySerAlaAlaLeuTyrGluSerTrpSerIleProPheSerValMetLeuValAla 905
Qy 1352 ATTCAATGACCTTTTGGCTACCTGCGAGTCTGCTTTATAGCAATTTCTTTGAAGAA 1411
Db 906 -----ProLeuGlyValIleGlyAlaLeuLeuAlaIleThrPhe-----ArgGly 920
Qy 1412 ATTCAATGATTTTGCAGCACTTGGCTTTATATGACTGATTTGGGTATTCATTAAAC 1471
Db 921 LeuThrAsnAspValTyrPheGlnValGlyLeuLeuThrThrIleGlyLeuSerAlaLys 940
Qy 1472 AATTCCTTGATCTTTTATGATCGTATTCGTA---GATGCCAAGCAACCTGTTT 1525
Db 941 AsnAlaIleLeuIleValGluPheAlaLysAspLeuMetAspLysGluGlyLysGlyLeu 960
Qy 1526 ACCCCATGATGATTTTATGATATGATGACCTTCGCAAAAGCGTTACGCCGACGATATG 1585
Db 961 IleGluAlaThrIleu-----AspAlaValArgMetArgLeuArgProIleLeuMet 977
Qy 1586 ACAACA---GCTCAACTATATGATGCTTTATATGCTTTGTTATAGCC---GGCTCC 1639
Db 978 ThrSerLeuAlaPheIleLeuGlyValMetProLeuValIleSerThrGlyAlaGlySer 997
Qy 1640 TCTGCTTAAATTTTGCATTATTAATGACCAAT---GGATTCCTTACGAACTTATTCG 1696
Db 998 GlyAlaGlnAsnAlaValAlaGlyThrGlyValMetGlyGlyMetValThrAlaThrValLeu 1017
Qy 1697 TCTCTTATATTCGACACCTCGTTGTTATATGCTTCGTTAAAGAAATTCG 1750
Db 1018 AlaIlePhePheValProValPhePheValValArgArgArgPheSerArg 1035

RESULT 21
US-09-815-242-5186
/ Sequence 5186, Application US/09815242
/ Patent No. US2002061569A1
/ GENERAL INFORMATION:
/ APPLICANT: Haselebeck, Robert
/ APPLICANT: Ohlsen, Karl L.
/ APPLICANT: Zyskind, Judith W.
/ APPLICANT: Wall, Daniel
/ APPLICANT: Twawick, John D.
/ APPLICANT: Carr, Grant J.
/ APPLICANT: Yamamoto, Robert T.
/ APPLICANT: Xu, H. Howard
/ TITLE OF INVENTION: Identification of Essential Genes in
/ TITLE OF INVENTION: Prokaryotes
/ FILE REFERENCE: ELITRA.011A
/ CURRENT APPLICATION NUMBER: US/09/815, 242
/ CURRENT FILING DATE: 2001-03-21
/ PRIOR APPLICATION NUMBER: 60/191, 078
/ PRIOR FILING DATE: 2000-03-21
/ PRIOR APPLICATION NUMBER: 60/206, 848
/ PRIOR FILING DATE: 2000-05-23
/ PRIOR APPLICATION NUMBER: 60/207, 727
/ PRIOR FILING DATE: 2000-05-26
/ PRIOR APPLICATION NUMBER: 60/242, 578
/ PRIOR FILING DATE: 2000-10-23
/ PRIOR APPLICATION NUMBER: 60/253, 625

```

[illegible]

Db	345		-----SerSerThrProLysThr	350
QY	603	GTAGTCAAGTTTAGAATCAAAATGGCT-----CCAAAAGCCTTGGTATATCCTTTTTCT	550	
Db	351	AlaValAlaLeuSerIalalysLysValTpArgProAlaAla	364	
QY	549	ACAGATTTTTTAAGACTTTTGAGACAATAAAAAATTCCTCTCGGATTCTTCGAATACAAGA	490	
Db	365	-----ArgSerThrSer	368	
QY	489	ACATTTGCATCTACGGCCATCCCATAGCAGNAGAACATCCACCGAGTCTCTGNCAAGGTG	430	
Db	369	TyrPheThrGlnThrProAlaValArgThrAspAlaProPheIleProAspHisVal	388	
QY	429	AGTGGCGCATCCAATACTGTACAGCTGCCAGATAAGCAAAGATTTCAGAGAACAACAGCT	370	
Db	389	GlnValAlaLysGlnHisValGlnSerIleAlaAlaPheGlnArg--argLysGlyLeu	407	
QY	369	CCGGAAGCGATGACGGCTCCAAATCTATAATATACGCTCATCAAAACAATAAGCATGGCC	310	
Db	408	ProArgGlyProThrAlaGluArgIle-----LeuAspSerAlaPheGlnValPro	424	
QY	309	AAGCCA-----CAGCATGCTGAGATAATGCCCTTGTCATACGTTTTTCCCAAGA	259	
Db	425	HisProCysHisPheGlnHisAlaAlaLeuProSerCysCysHisCysPheProArg	444	
QY	258	TCAGAAAGATCGCTCTTTCACCTGAGNACTCGGGAAACAAAGACATCCGTCACAGATTTT	199	
Db	445	Cys-LysLysAsnProPheTyrrAlaProGlnIleGln---ThrAsnLeuProPheTh	463	
QY	198	AAATCTGACGGCGAGTTTGCTCACTTCACGGTGGGTAATTTTCCTCGAGCACACTGGGCATGA	139	
Db	463	rProAlaLysGlu-----SerLeuValHisTrpMetse	474	
QY	138	TTTTTCAATGGGACGTTTAAATAGGGCTGCTGACCATATAAC	96	
Db	474	rSerSerCysArgCysLeu-----CysAlaProArqAsn	485	

```

RESULT 24
US-09-815-242-10968
; Sequence 10968, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: EPIFRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 10968

```

LENGTH: 396
 TYPE: PRT
 ORGANISM: Haemophilus influenzae
 US-09-815-242-10968

Alignment Scores:

Pred. No.:	0.00637	Length:	396
Score:	109.00	Matches:	64
Percent Similarity:	33.43%	Conservative:	52
Best Local Similarity:	18.44%	Mismatches:	97
Query Match:	3.29%	Indels:	134
	10	Gaps:	14

US-09-868-987-1 (1-1864) x US-09-815-242-10968 (1-396)

```

QY 1037 TTGGATCTTCAGAAAGATCAAAATC--TATTGATGATTAAGCTTAAGCTTAAGTACT 1093
Db 25 PheasnThrThrGluPheValProValAlaMetLeuSerAspIleAlaGlnSerPheasp 44
QY 1094 AAGCAGATACGAGCCTCTCTCTAAATTAACGATCATGACGTCGCTTATGTGGATT 1153
Db 45 MetGlnThrAlaSerThrGlyLeuMetMetThrValTyralatPrThr-----Val 61
QY 1154 GTTGTCGAAACAGCCCAAGTTTCTC--TACGAAACCTCTAAACGA----- 1198
Db 62 LeuIleMetSerLeuProAlaMetLeuAlaThrGlyAsnMetGluArgGlySerLeuLeu 81
QY 1198 ----- 1198
Db 82 IleIysLeuPheIleIlePheIleValGlyHisIleLeuSerValIleAlaTrpAsnPhc 101
QY 1199 -----AACGAAATTTGGTCAAG 1219
Db 102 TrpIleLeuLeuAlaArgMetCysIleAlaLeuAlaHisSerValPheTrpSerIle 121
QY 1220 GTAGCAGC-----AACTATCGAAGAAATGCTTATCGAGCAGCATCGGCTT 1270
Db 122 ThrIaSerLeuValMetArgIleSerProIshIshIshIshIshIshIshIshIshIsh 141
QY 1271 TTAGCA-----GCTTGGCAATCATCTTG-----CTCTATGTGAGT 1306
Db 142 LeuAlaIleGlyThrAlaLeuAlaThrIleLeuGlyLeuProIleGlyArgIleValGly 161
QY 1307 TTGGCTTTGAATGCAATATGCTTTCAGTCCGATGCGCTTATTCATGACCTTTTG 1366
Db 162 GlnLeuValGlyTrpArgValThrPheGlyIleIleAlaValLeuAlaLeuSerIleMet 181
QY 1366 ----- 1366
Db 182 PheLeuIleIleArgLeuLeuProAsnLeuProSerIshAsnAlaGlySerIleAlaSer 201
QY 1367 -----GCTACCTGTGACGCTTG 1384
Db 202 LeuProIleLeuAlaIshArgProIleLeuLeuTrpLeuTyValThrThrAlaIleVal 221
QY 1385 TTATAGACATTTTC-----TTTTGAAAGAAATTCATATGATTTGCAAGCC 1432
Db 222 IleSerAlaHisPheThrAlaTyThrTyrlleGluProPheMetIleAspValGlyHis 241
QY 1433 ATTTGT-----GCTTATGACTGTGATGGGTATGCTTAAACATATCT 1477
Db 242 LeuAspProAsnPheAlaThrAlaValLeuLeuValPheGlyPheSerGlyIleAlaAla 261
QY 1478 TTGATCATTTTGAATCGTATTCGTAAGATGCGCAAGCAACGTTTACCCCTATGAT 1537
Db 262 SerLeuLeuPheAsnArgLeuTyArg-----PheAlaProThrIsh 275
QY 1538 GTTTAGTATGATGCTTCCTCAAAAGAGCTTACGCGCAGCGATGATGACAACAGCTACA 1597
Db 276 PheIleValIshSerMetSerLeuLeuMetPheSerLeuLeuLeuLeuPheSerThr 295
QY 1598 ---ACTCATATAGTTTGTATATGCTTTTGTATTAGCGCGC----- 1636

```

```

Db 296 LysThrIleIleAlaMetPheSerLeuValPheIleTrpGlyIleGlyIleSerCysIle 315
QY 1636 ----- 1636
Db 316 GlyLeuSerLeuGlnMetArgValLeuIshLeuAlaProAspAlaThrAspValAlaThr 335
QY 1637 -----TCCCTGTCCTTATTTTGGATTTATATGACCATGAGGATTTCTTAGGA 1687
Db 336 AlaIleTySerIshIlePheAsnAlaGly-----IleGlyAlaGlyAlaLeuPheGly 353
QY 1688 ACTTTATGCTCTCTTATAT 1708
Db 354 AsnLeuAlaThrThrTyrlleu 360

```

RESULT 25

```

US-09-815-242-5331
/ Sequence 5331, Application US/09815242
/ Patent No. US20020061569A1
/ GENERAL INFORMATION:
/ APPLICANT: Haselbeck, Robert
/ APPLICANT: Ohlsen, Kari L.
/ APPLICANT: Zyskind, Judith W.
/ APPLICANT: Wall, Daniel
/ APPLICANT: Trawick, John D.
/ APPLICANT: Carr, Grant J.
/ APPLICANT: Yamamoto, Robert T.
/ APPLICANT: Xu, H. Howard
/ TITLE OF INVENTION: Identification of Essential Genes in
/ FILE REFERENCE: ELITRA.011A
/ CURRENT APPLICATION NUMBER: 2001-03-21
/ PRIOR FILING DATE: 2000-09/815,242
/ PRIOR APPLICATION NUMBER: 60/191,078
/ PRIOR FILING DATE: 2000-03-21
/ PRIOR APPLICATION NUMBER: 60/206,848
/ PRIOR FILING DATE: 2000-05-23
/ PRIOR APPLICATION NUMBER: 60/207,727
/ PRIOR FILING DATE: 2000-05-26
/ PRIOR APPLICATION NUMBER: 60/242,578
/ PRIOR FILING DATE: 2000-10-23
/ PRIOR APPLICATION NUMBER: 60/253,625
/ PRIOR FILING DATE: 2000-11-27
/ PRIOR APPLICATION NUMBER: 60/257,931
/ PRIOR FILING DATE: 2000-12-22
/ PRIOR APPLICATION NUMBER: 60/269,308
/ PRIOR FILING DATE: 2001-02-16
/ NUMBER OF SEQ ID NOS: 14110
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 5331
/ LENGTH: 651
/ TYPE: PRT
/ ORGANISM: Staphylococcus aureus
US-09-815-242-5331

```

Alignment Scores:

Pred. No.:	0.00993	Length:	651
Score:	108.00	Matches:	110
Percent Similarity:	35.62%	Conservative:	87
Best Local Similarity:	19.89%	Mismatches:	202
Query Match:	3.26%	Indels:	154
	10	Gaps:	26

US-09-868-987-1 (1-1864) x US-09-815-242-5331 (1-651)

```

QY 305 GGCTTGCAATGCTTATTTGATGACGCTATAT-----TATGATTTGGA 352
Db 42 GlyLeuMetIleIlePheProAlaIleIleGlyTyrlleuGlyGlnPheGlnPheGly 61
QY 353 GCGCTCATGCTTCGGAGCGTGTCTTGATCTTGTATCTGAGGAGCTTACAG 412
Db 62 LeuLeuValAlaThrGlyThrLeu-----AlaHisIleTy 73
QY 413 TATTGATGCGCA-----CTACCTTGTGAGAGCTGCGT--- 448

```

Db 74 ValPheLysGlyProSerArgSerLysLeuArgThrValIleIleCysAsnLeuAlaPhe 93
Qy 449 GGGATTGTTGCTGATGGGATGCCCGTAGATGCAATGTTCTTGATTCGAAAGATC 508
Db 94 AlaIleCysMetMetLeuGlyThrIleuThrAlaLysThrProLeuValPhe----- 110
Qy 509 CGAGAGGAATTTTATTGCTCTCAAGCTTAAAGCTTAAAGAAATCTGAGAAAAAGGATATACCAAG 568
Db 111 -----GlyMetThrLeu 114
Qy 569 GCTTTTGAGCCATTTTGAATCTTAACCTTGACTACAGTATTGGCCTCAGCA----- 619
Db 115 LeuIleValThrValIleProPheThrIlePheThrAlaLeuLysIleAlaGlyProSer 134
Qy 620 CTTCTTTTCTCTAGATACAGGG-----CCTATT----- 649
Db 135 SerThrPhePheIleValThrPheSerLeuProIleAsnLeuProIleAlaProGluGlu 154
Qy 650 -----AAGGGTTGCTTGTACATGATTTTAGAAATTTCTCTCTTCAATGTTTACG 700
Db 155 AlaLeuTyArgGlyPheAlaIle---LeuValGlyGlyIleLeuAlaThrMet----- 171
Qy 701 GCTCTTTTCATGACTAAATTTTCTTCATGCTGTGATGAATAGACCAACATACACAG 760
Db 172 -----MetValLeuIleThrIleValPheSerLysAsnLysAlaGluGluGlnAla 188
Qy 761 TTGCAT-----ATGATGAATAAGTTGCGGGGATAAAGCATGAT----- 799
Db 189 IleGlnAsnAspPheLysLeuIleSerLysLeuLeuHisThrTyAsnAspLysSerAla 208
Qy 800 TTTCTGAGAGGATGCAAA-----AAACTTTGGGCT 829
Db 209 PheLeuLysValAlaLysThrAlaValAspSerPheLysAlaSerAspLysLeuLeuIle 228
Qy 830 GTTCTGGAAGTGTTCCTTTAGTTGCGTGTCTCGGGTTGGAGCCGGAATCC 889
Db 229 ThrSerThrSerSer-----AsnAsp 235
Qy 890 GTTTTGGGAATGGAATTTAAAGAGGGTATGCTTTTACCTTTAAATCCAAAGAGCATGGC 949
Db 236 LysLeuSerArgArgPheGln-----LysLeuLeuLeuLeuHisThrSerAlaGlnGly 253
Qy 950 ATC---ACGAGTGTGCTCAATGCGTGGCAAGTTGTGCATAAATACAGGAAGCTGCT 1006
Db 254 IleTySerGluLeuGluLeuAsnAlaLysGlnIleArgProLeuProAspGluLeu 273
Qy 1007 CTTTCTCTAGAGCTTCGGTATCAACATTTGGATCTTCAGAAAAGATCAAAATCTAT 1066
Db 274 IleGluMetMetAspHisIleIleAlaGlnLeuAspAsnSerAspGlu---AsnValArg 292
Qy 1067 TTTAGTGATAAAGCTTTAAGCTATATAAGCAGATACGA-----GCCTCTCTCCTA 1117
Db 293 TyrTrpArgLysGluValThrValThrGluGluPheGlnAsnLeuPheAsnHisIleLeu 312
Qy 1118 AAATTAACGATCATGAGCTGGCGTTATTGTGGGATTTGTGAGAAACAGCGCTAGATTT 1177
Db 313 LysIleAspGluMet-----ValHisAlaAsnGluAlaArgIle 325
Qy 1178 CTCTACGGAACCTTAACGAAACCCAAATTTTGGTCAAGGTAAGCAGCAAACTATCG 1237
Db 326 AlaTyGluAlaAspMetArgLysProLeuTySerLysArgIleTyGlnAsnLeuThr 345
Qy 1238 -----AAGAAATGCGTTATCAGGCGCATCGGGCTTTTAGGA 1276
Db 346 LeuAspSerIleValPheArgAsnThrLeuArgIleThr----- 358
Qy 1277 GCTTTGGCAATCATCTGCTATGTAGTTTGGCTTTGAATGCAATATGCTTTTCACT 1336
Db 359 AlaIleMetMetIleAlaIlePheIleAlaLeuMetPheAsnPheGluLysAlaTyTrp 378
Qy 1337 GCCGTATCGGCTTTAATTCATGACCTTTTGGCTACTGCTGCACTGTTGTTATACCAT 1396

Db 379 IleProLeuSerAlaHisThrIleLeuLeuGlyThrSerThrIleHisAlaIleGluArg 398
Qy 1397 TTTCTTTTGAAGAAATTTCAAATAGATTTTCAAGCCATTGGTGTCTTAATGACTGTATTG 1456
Db 399 GlyMetAlaArg-----GlyLeuGlyThrIleLeu 408
Qy 1457 GGGTATTTCATTAAACAATACCTTTCATCATTTTCAT-----CGTATTTCGT 1501
Db 409 GlyValLeuValLeuSerValIleLeuLeuPheSerIleProThrProValAlaValIle 428
Qy 1502 GAAGATCGCCCAAGGAACCTGTTTACCCCTATGATCATGTTTGTAGTTAATGATGCTTCA 1561
Db 429 LeuMetGlyIleAlaAlaLeuPheThrGluAlaLeuValGlyAlaAsnTyAlaIleAla 448
Qy 1562 AAGAGCTTTCAGCCGCGGTAATGACAACAGCTACAACCTATCATGATTTTGTATGCTT 1621
Db 449 ValValPhe-----IleThrIleGlnValIleLeuMetAsn 460
Qy 1622 TTGTTTATAGCGGCTCTCTGCTCTTAAATTTTCATTT----- 1660
Db 461 GlyLeuAlaSerGlnAsnLeuThrIleAsnIleAlaPheProArgValIleAspValAla 480
Qy 1661 -----ATTATGACCATAGGATTT---CTTCTAGGAACCTTTATCGTCTCTCTTAT 1705
Db 481 MetGlyIleValIleAlaIleIleGlyLeuPheValLeuGlyGlnArgThrAlaSerAla 500
Qy 1706 ATTCACACACCTCTGTTGTTGTTTATGTCGCTAAAGAA 1744
Db 501 LeuLeuProAsnValMetAlaGluValValArgLysGlu 513
RESULT 26
US-09-815-242-12682
; Sequence 12682, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12682
; LENGTH: 651
; TYPE: PRN
; ORGANISM: Staphylococcus aureus
US-09-815-242-12682
Alignment Scores: 0.00993 Length: 651
Pred. No.: 108.00 Matches: 110
Score:

Percent Similarity:	35.62%	Conservative:	87
Best Local Similarity:	19.89%	Mismatches:	202
Query Match:	3.26%	Indels:	154
DB:	10	Gaps:	26

Oy	305	GGCTTGGCAATGCTTATGTTTGGATGAGCGTATAT-----TATAGATTGGA	355
Db	42	GlyLeuLeuMetIleIleIleProAlaIleIleIleTyrlLeuCySglYAsnPhnGlnPhnGly	61
Oy	353	GGCGTCATCGCTTCGGAGCGTGTCTTCTGAACTTTTGTCTTATCTGGAGCGCTACAG	412
Db	62	LeuLeuValAlaThrGlyThrLeu-----AlaHisIleTyr	73
Oy	413	TATTTGGATGGCCCA-----CTACACTGTGAGACTCGCT--	448
Db	74	ValPhenylSglYProSerArgSerIleuArgThrValIleIleCyAsnLeuAlaPhe	93
Oy	449	GGGATGTTCTTGTGCATAGGGAGCGCCGTAGATCAAAATGTTCTTATTCGAAAGATC	508
Db	94	AlaIleCyMetMetLeuGlyThrLeuThrAlaLysThrProLeuValPhe-----	110
Oy	509	CGAGAGAAATTTTATTTCTCCAAAGCTTTAAAAATCTGAGAAAAAGATATACCAG	568
Db	111	-----GlyMetThrLeu	114
Oy	569	GCTTTTGAGCAATTTTGATTTCTATCTACATTCAGTACAGATTGGCCTCAGA-----	619
Db	115	LeuIleValThrValIleProPheTyrlPheThrAlaLeuLysIleAlaGlyProSer	134
Oy	620	CTTCTTTTCTTCTCTAGATACAGG-----CCTATT-----	649
Db	135	SerThrPhePheIleValThrPheSerLeuProIleAsnLeuProIleAlaProGlu	154
Oy	650	-----AAAGGTTGCTTTGATATTTAGAGATTTTCTCTCATCTTTTACG	700
Db	155	AlaLeuTyrlArgGlyPheAlaIle--LeuValGlyGlyIleLeuAlaThrMet-----	171
Oy	701	GCTCTTTTCATGACTAAATTTTCTTCATGCTGTGGATGAATAAGCCCAATACAG	760
Db	172	-----MetValLeuIleThrIleValPheSerLysAsnLysAlaGluGlnAla	188
Oy	761	TTTGCA-----ATGATGAATAGTTCGTGGGGAATAAGCAATGAT-----	799
Db	189	IleGlnAsnAspPheLysLeuIleSerIleuLeuHisIleThrTyrlAsnAspLysSerAla	208
Oy	800	TTCTTTGAGAGATGCANA-----AAACTTTGGGCT	829
Db	209	PheLeuLysValAlaLysThrAlaValAspSerPheLysAlaSerAspLysLeuIle	228
Oy	830	GTTTCTGGAAGTGTTTTCTTTAGGTTCGTTGCTCTCGGAGTTTGGAGCTGGAAATCC	889
Db	229	ThrSerThrSerSer-----AsnAsp	235
Oy	890	GTTTGGGGAATGATTTTAAAGAGGGATGCTTTACTTTAATCCAAAGAGCATGGC	949
Db	236	LysLeuSerAspArgArgPheGln-----LysLeuLeuLeuLeuHisIleThrSerAlaGlnGly	253
Oy	950	ATC-----AGCGATGTGCTCAAAATCGTGGCAAAAGTTGTGCATAACTACAGAAAGCTGT	1000
Db	254	IleTyrlSerGluLeuLeuGluIleuLysAsnAlaLysGlnIleArgProLeuProAspGluLeu	273
Oy	1007	CTTTCTTCAAGACCTTCGATTCGATTAACAATTTGGATCTTCAGAAAAAGATCAAAATCTAT	1066
Db	274	IleGluMetMetAspHisIleIleIleAlaGlnLeuAspAsnSerAspGlu-----AsnValArg	292
Oy	1067	TTTATGATTAAGCTTTTAAAGCTATGATCTACAGAGATACAG-----GCGCTCTCTCTA	1111
Db	293	TyrlThrArgLysGluValThrValThrGluIleuPheGlnAsnIleIleLeu	312
Oy	1118	AAATTAAGCATATGACGCGCGCTTATTTGTGGAGATTGTTCGCAAAAACAGGCGCTAGATT	117

Db	313	Lys1leAsp1IuMet-----	-----ValH1AlAsnGlu1AlaArg1le	325
Qy	1178	CTTACGGAACCTCTTAAGCAACGAAATTTGGTCAAGGTAAAGCAAACTATCG	1237	
Db	326	Ala1yGlu1AlaAspMetArgLysProLeuTySerLysArg1IeTyGlnAsnLeuThr	345	
Qy	1238	-----AAGAAATAGCGCTTATCAGCGACACATCGGGCTTTAGCA	1276	
Db	346	LeuAspSer1IeValPheArgAsnThrLeuAlaGlyTrpThr-----	358	
Qy	1277	GCTTTGGCAATCATCTTGGCTCTATGTAGAGTTTGCCTTGAATGGCAATATGCTTTCAGT	1336	
Db	359	Ala1IeMetMet1IeAla1IePhe1IeAla1IeLeuMetPheAsnPheGluLysAlaTrp	378	
Qy	1337	GCGGTATGCCCTTAAATCATGACCTTTGGCTACCTGTGACAGCTCTTGTATATACACAT	1398	
Db	379	1IeProLeuSer1Ala1Thr1IeLeuGlyThrSerThr1IeH1Ala1IeGluArg	398	
Qy	1397	TTCTTTTGGAGAAATTCAAATATGATTTTGCAAGCATTTGGCTTATGACGTGATTG	1456	
Db	399	GlyMetAlaArg-----GlyLeuGlyThr1IeLeu	408	
Qy	1457	GGGTATTCATTTAAACATATCTTGTATCTTTTGAT-----CGTATTCGT	1501	
Db	409	GlyValLeuValLeuSerVal1IeLeuLeuPheSer1IeProThrProValAlaVal1Ie	428	
Qy	1502	GAGATGCGCCAGACGAACTGTTTACCCTCATGTATGTTTATGTTAATGATGCCCTTCAA	1561	
Db	429	LeuMetGly1IeAla1AlaLeuPheThrGluAlaLeuValGlyAlaAsnTyx1Ala1IeAla	448	
Qy	1562	AAGCGTTGACCGCGACGGTAATGACACAGCTACATCTATGAGTTTGTATATGCTT	1621	
Db	449	ValValPhe-----1IeThr1IeGlnVal1IeLeuMetAsn	460	
Qy	1622	TTGTTTATAGCGCGCTCTCTCTGTCTTAAATTTTGCAATT-----	1660	
Db	461	GlyLeuAlaSerGlnAsnLeuThr1IeAsn1IeAlaPheProArgVal1IeAspValAla	480	
Qy	1661	-----ATTTATGACCATATGAGGATTT-----CTTTAGGAACCTTATCGTCTCTTAT	1705	
Db	481	MetGly1IeVal1IeAla1IeAla1IeGlyLeuPheValLeuGlyGlnArgThrAlaSerAla	500	
Qy	1706	ATTGACACCTCGTTTGTGTTTATGATGCCGAAAGAA	1744	
Db	501	LeuLeuProAsnValMetAlaGluValAlaArgLysGlu	513	
RESULT 27				
US-09-815-242-13719				
Sequence 13719, Application US/09815242				
Patent No. US20020061569A1				
GENERAL INFORMATION:				
APPLICANT: Haselbeck, Robert				
APPLICANT: Ohlsen, Karl L.				
APPLICANT: Zyskind, Judith W.				
APPLICANT: Wall, Daniel				
APPLICANT: Trawick, John D.				
APPLICANT: Carr, Grant J.				
APPLICANT: Yamamoto, Robert T.				
APPLICANT: Xu, H. Howard				
TITLE OF INVENTION: Identification of Essential Genes in				
TITLE OR INVENTION: Prokaryotes				
FILE REFERENCE: ELITRA.011A				
CURRENT APPLICATION NUMBER: US/09/815,242				
CURRENT FILING DATE: 2001-03-21				
PRIOR APPLICATION NUMBER: 60/191,078				
PRIOR FILING DATE: 2000-03-21				
PRIOR APPLICATION NUMBER: 60/206,848				
PRIOR FILING DATE: 2000-05-23				
PRIOR APPLICATION NUMBER: 60/207,727				
PRIOR FILING DATE: 2000-05-26				
PRIOR APPLICATION NUMBER: 60/242,578				
PRIOR FILING DATE: 2000-10-23				
PRIOR APPLICATION NUMBER: 60/253,625				

; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 13719
; LENGTH: 1049
; TYPE: PRN
; ORGANISM: Salmonella typhi
US-09-815-242-13719

Alignment Scores:
Pred. No.: 0.0122 Length: 1049
Score: 108.00 Matches: 122
Percent Similarity: 31.62% Conservative: 100
Best Local Similarity: 17.38% Mismatches: 240
Query Match: 3.26% Indels: 240
DB: 1.0 Gaps: 29

US-09-868-987-1 (1-1864) x US-09-815-242-13719 (1-1049)

QY 308 TTGCAATGCTTATTTGTTGATGAGCGTATATATAGATTGGAGGCGTCATCGCTTCG 367
Db ||||| : : : : : |||||
QY 353 LeuValMetTyrLeuPheLeuGlnAsnPheArgAlaThrLeuIleProThrIleAlaVal 372
Db ||||| : : : : : |||||
QY 368 GGAGCTGTTCTTCTGAATCTTTGCTTATCTGGGAGCTCTACAGATTATGGATCGGCCA 427
Db ||||| : : : : : |||||
QY 373 ProValValLeuLeuGlyThrPheAlaValLeuAlaPheGlyPheSerIleAsnThr 392
Db ||||| : : : : : |||||
QY 428 CTCACCTTGTGAGGACTCGCTGGGATGTTCTTGTATGGGATGGCGTATAGTCAAAAT 487
Db ||||| : : : : : |||||
QY 393 LeuThrMetPheGlyMet-----ValLeuAlaIleGlyLeuValAspAla 409
Db ||||| : : : : : |||||
QY 488 GTTCTTGTATTCGAAGAATCCGAGAGAAATTTTATGTCTCAAGCTTT-----AAA 541
Db ||||| : : : : : |||||
QY 410 IleValValGluAsnVal---GluArgValMetThrGluGluGlyLeuProProlys 428
Db ||||| : : : : : |||||
QY 542 AAATCTGTAGAAAAGGATATACCAAGGCTTTTGGAGCCATTTTGTATCTAACTTGACT 601
Db ||||| : : : : : |||||
QY 429 GluAlaThrArgLysSerMetGlyGlnIleGlnGlyAlaLeuVal-----GlyIleAla 446
Db ||||| : : : : : |||||
QY 602 ACAGTATGCTCTACGACTTCTT-----TTCTTCTAGATACAGGCGCTATT 649
Db ||||| : : : : : |||||
QY 447 MetValLeuSerAlaValPheIleProMetAlaPhePheGlyGlySerThrGlyAlaIle 466
Db ||||| : : : : : |||||
QY 650 ---AAAGGTTTGTCTTACATTTAGTATTTAGGAATTTCTTCAATGTTTACGCTCTT 706
Db ||||| : : : : : |||||
QY 467 TyrArgGlnPheSerIleThrIleValSerAlaMetAlaLeuSerValLeuAlaLeu 486
Db ||||| : : : : : |||||
QY 707 TTCATGACT-----715
Db ||||| : : : : : |||||
QY 487 IleLeuThrProAlaLeuCysAlaThrMetLeuLysProValAlaLysGlyAspHisGly 506
Db ||||| : : : : : |||||
QY 716 -----AAATTTTCTCATGCTGTGGATGAATAAGACCCACATACAGTGTGAT--- 766
Db ||||| : : : : : |||||
QY 507 GluGlyLysGlyPhePheGlyTrpPheAsnArgLeuPheAspLysSerThrHisHis 526
Db ||||| : : : : : |||||
QY 766 -----766
Db ||||| : : : : : |||||
QY 527 TyrThrAspSerValTrpAsnIleLeuArgSerThrGlyArgTyrLeuLeuLeuTyrLeu 546
Db ||||| : : : : : |||||
QY 767 -----ATGATGAATAAGTTCGTGGGATAAAGCATGATTCTTGT----- 805
Db ||||| : : : : : |||||
QY 547 IleIleValValGlyMetAlaTyrLeuPheValArgLeuProSerSerPheLeuProAsp 566
Db ||||| : : : : : |||||
QY 805 -----805
Db ||||| : : : : : |||||
QY 567 GluAspGlnGlyValPheLeuThrMetValGlnLeuProAlaGlyAlaThrGlnGluArg 586
Db ||||| : : : : : |||||
QY 806 -----AGAGGATGC 814
Db ||||| : : : : : |||||

Db 587 ThrGlnLysValLeuAspGluValThrAspTyrTyrLeuAsnLysGluLysAlaAsnVal 606
QY 815 AAAAATCTTTGGGCTGTTTCTGGAAGTGTCTTTCTT-----TTAGGTTC 859
Db ||||| : : : : : |||||
QY 607 GluSerValPheAlaValAsnGlyPheGlyPheAlaGlyArgGlyGlnAsnThrGlyIle 626
Db ||||| : : : : : |||||
QY 860 GTTCTCTCGGTTTGGAGCTCG-----AATTCGTTTGGGA 898
Db ||||| : : : : : |||||
QY 627 AlaPheValSerLeuLysAspTrpAlaAspArgProGlyGluLysAsnLysValGluAla 646
Db ||||| : : : : : |||||
QY 899 ATG-----GATTTTAAAGGAGGATGCTTACCTTTACCTTT 931
Db ||||| : : : : : |||||
QY 647 IleThrGlnArgAlaThrAlaAlaPheSerGlnIleLysAspAlaMetValPheAlaPhe 666
Db ||||| : : : : : |||||
QY 932 AAT-----934
Db ||||| : : : : : |||||
QY 667 AsnLeuProAlaIleValGluLeuGlyThrAlaThrGlyPheAspPheGluLeuIleAsp 686
Db ||||| : : : : : |||||
QY 934 -----934
Db ||||| : : : : : |||||
QY 687 GlnAlaGlyLeuGlyHisGluLysLeuThrGlnAlaArgAsnGlnLeuPheGlyGluVal 706
QY 935 -----CCAAAAGAGCATGCGCATCAGCGATGTTGCTCAA 967
Db ||||| : : : : : |||||
QY 707 AlaLysTyrProAspLeuValGlyValArgProAsnGlyLeuGluAspThrProGln 726
QY 968 ATGCTGCGCAAGTT---GTGCATAAACTACAGAAAGCTGCTTCTTCTTAGAGACTTC 1024
Db ||||| : : : : : |||||
QY 727 PheLysIleAspIleAspGlnGluLysAlaGlnAlaLeuGlyValSerIleSerAspIle 746
Db ||||| : : : : : |||||
QY 1025 CGTATT-----CAAAATTTGGATCTTCAAAAAGATCAAAATCTATTATTAGTGAT 1075
Db ||||| : : : : : |||||
QY 747 AsnThrThrLeuGlyAlaAlaTrpGlySer-----TyrValAsnAsp 761
Db ||||| : : : : : |||||
QY 1076 -----AAAGCTTTAAGCTATATAAGCAGATACGAGCTCTCTC 1114
Db ||||| : : : : : |||||
QY 762 PheIleAspArgGlyArgValLysValTyrValMetSerGluAlaLysTyrArgMet 781
QY 1115 CTAATAATTAACGATCATGCTGGCTTATTGTGGGATTGTTGTCAAGAACAGGCT--- 1171
Db ||||| : : : : : |||||
QY 782 LeuProAspAspIleAsnAspTrpTyrValArgGlySerAspGlyGlnMetValProPhe 801
Db ||||| : : : : : |||||
QY 1172 -----AGATTTCTCTACGAACTCTCAA--- 1195
Db ||||| : : : : : |||||
QY 802 SerAlaPheSerSerArgTrpGluTyrGlySerProArgLeuGluArgTyrAsnGly 821
Db ||||| : : : : : |||||
QY 1196 -----CGAAACGCA 1204
Db ||||| : : : : : |||||
QY 822 LeuProSerMetGluIleLeuGlyGlnAlaAlaProGlyLysSerThrGlyGluAlaMet 841
QY 1205 AAATTTTGGTCAAGGTAAGCAGCAAACTATCGAAG-----AAA 1243
Db ||||| : : : : : |||||
QY 842 AlaMetMetGluGluLeuAlaSerLysLeuProSerGlyIleGlyTyrAspTrpThrGly 861
QY 1244 ATGCTTATACAGGCGCATCTCGG-----CTTTTAGGAGCTTTGGCAATC 1288
Db ||||| : : : : : |||||
QY 862 MetSerTyrGlnGluArgLeuSerGlyAsnGlnAlaProAlaLeuTyrAlaIleSerLeu 881
Db ||||| : : : : : |||||
QY 1289 ATCTTGCTCTATGTG-----AGTTTGGCTTTGATCGCAATATGCTTTCAGTCC 1339
Db ||||| : : : : : |||||
QY 882 IleValValPheLeuCysLeuAlaLeuTyrGluSerTrpSerIleProPheSerVal 901
QY 1340 GTATGCGCTTAAATTCATGACCTTTTGCTACCTGTCAGCTCTGTTTATAGCACATTTC 1399
Db ||||| : : : : : |||||
QY 902 MetLeuValVal-----ProLeuGlyValIleGlyAlaLeuLeuAlaAlaThrPhe 918
Db ||||| : : : : : |||||
QY 1400 TTTTGAAGAAAATTCAAATAGATTTCAGCCCATTTGGCTTTTAAGTACTGTATTGGGG 1459
Db ||||| : : : : : |||||
QY 919 -----ArgGlyLeuThrAsnAspValTyrPheGlnValGlyLeuLeuThrThrIleGly 936
Db ||||| : : : : : |||||
QY 1460 TATTCAATTAACAATACTTTCATCATTTTTCATCGTATTTCGTGAA-----GATCGCCAA 1513
Db ||||| : : : : : |||||
QY 937 LeuSerAlaLysAsnAlaIleLeuIleValGluPheAlaLysAspLeuMetAspLysGlu 956
Db ||||| : : : : : |||||

QY 1514 GCCAACCTGTTTACCCCTATGATGTTTATGATGATGCTTCAAAAGACCTTGAC 1573
 Db 957 G1yLeuGlyLeuValGluAlaMetLeu-----GluAlaValArgMetArgLeuArg 973
 QY 1574 CGCAGCGTATGCAACA-----GCTAACACTCTATGCTTTGTTAAGCTTTGTTATA 1630
 Db 974 ProIleuMetThrSerLeuAlaPheMetLeuGlyValMetProLeuValIleSerSer 993
 QY 1631 GGC---GGCTCTCTGCTCTTATTTGATTTATGACATTA---GGGATCTTCTTA 1684
 Db 994 G1yAlaGlySerGlyAlaGlnAsnAlaValGlyThrGlyValLeuGlyGlyMetValThr 1013
 QY 1685 GGAATTATGCTCTCTTATTTATGACCACTCTGTTGTTGTTATGTCCTGTAAGAA 1744
 Db 1014 AlaThrValLeuAlaIlePhePheValProValPhePheValValArgArgArgPhe 1033
 QY 1745 AATGCC 1750
 Db 1034 SerArg 1035
 RESULT 28
 US-10-094-059-4
 / Sequence 4, Application US/10094059
 / Patent No. US20020127650A1
 / GENERAL INFORMATION:
 / APPLICANT: Curtis, Rory P.J.
 / TITLE OF INVENTION: 32466, A Human Sugar Transporter Family Member and
 / TITLE OF INVENTION: Uses Therefor
 / FILE REFERENCE: MP101-04CPIRM
 / CURRENT APPLICATION NUMBER: US/10/094,059
 / PRIOR FILING DATE: 2002-03-08
 / PRIOR FILING DATE: 2001-C3-12
 / NUMBER OF SEQ ID NOS: 8
 / SOFTWARE: FastSeq for Windows Version 4.0
 / SEQ ID NO 4
 / LENGTH: 488
 / TYPE: PRF
 / ORGANISM: Artificial Sequence
 / FEATURE:
 / OTHER INFORMATION: consensus
 / US-10-094-059-4
 Alignment Scores:
 Pred. No.: 0.0141 Length: 488
 Score: 106.00 Matches: 108
 Percent Similarity: 36.00% Conservative: 72
 Best Local Similarity: 21.60% Mismatches: 167
 Query Match: 3.20% Indels: 153
 DB: 12 Gaps: 20
 US-09-868-987-1 (1-1864) x US-10-094-059-4 (1-488)
 QY 596 TTGACTACAGTATTGGCTTCAGACTTTCTTCTTCTTCAATACAGGGCTATTAAAGG 655
 Db 3 LeuValAlaIleLeuGlyGlyPheLeuPheGlyTyrAspThrGlyValIleGlyGly 22
 QY 656 TTT-----GCTTTG 664
 Db 23 PheLeuAlaLeuIleAspPheLeuPheArgPheGlyLeuLeuThrSerSerGlyAlaLeu 42
 QY 665 ACATTGATTTTAAAGATTTCTCTTCAATGTTTACGGCTTTTCAAGACTTAATTTTTC 724
 Db 43 AlaGluLeuValGly---TyrSerThrValLeuThrGlyLeuValValSerIlePhePhe 61
 QY 725 TTCATGCTGTGATGAATAGACCAACATACACAGTTGATGATGATGAATAGTTGCTG 784
 Db 62 LeuGlyArgLeuIleGly-----SerLeuPheAla 71
 QY 785 GGGATTAAGCATGATTTCTTGAAGAGATGCAAAAACCTTGGGCTGTTTCTGGAAGTGT 844
 Db 785 GGGATTAAGCATGATTTCTTGAAGAGATGCAAAAACCTTGGGCTGTTTCTGGAAGTGT 844

Db 72 GlyLeuGlyAspArgPhe---GlyArgGlySerSerLeuLeuIleAlaLeuValLeu 90
 QY 845 TTTCTTTTAAAGTTGCGTTGCTCTCGGGTTTGAAGCTTGA-----ATTCCGTTTGG 895
 Db 91 PheValIleGly-AlaLeuLeuSerGlyAlaIleProGlyTyrThrIleGlyLeuThr 110
 QY 896 GGAATGGAATTTTA----- 908
 Db 110 PheAlaPheTyrLeuLeuIleValGlyArgValLeuValGlyLeuGlyValGlyAlaLeu 130
 QY 909 -----AAGAGGGTATGCTT 925
 Db 130 rValLeuValProMetTyrIleSerGluIleAlaProlyAlaLeuArgGlyAlaLeuGly 150
 QY 926 ACCTTTAATCCAAAAGAGATGAGCATGACGATGTTGCTCAATGCTGGCAAAAGTTTG 985
 Db 150 ySerLeuTyrGlnLeuAlaIleThrIleGlyIleLeuValAlaAlaIleIleGlyLeuGly 170
 QY 986 CATTAACCTACAGAGAG----- 1001
 Db 170 yLeuAsnLysThrAsnAsnAspSerAlaLeuAsnSerTyrGlyTyrArgIleProLeuGly 190
 QY 1002 -----CTGGCTTTCTTCTTGAAGACTTCCGTTATTCAA 1033
 Db 190 yLeuGlnLeuValProAlaLeuLeuLeuLeuIleGlyLeuLeu-PheLeuProGlySerP 210
 QY 1034 ACATTGGATCTTCAGAAAAGATCAATCTATTTATGATGATTAAGCTTTAAGCTATPACT 1093
 Db 210 roArgTyrLeuVal-GluLysGlyLysLeuGluGluAlaArgGluValLeuAlaLysLeu 229
 QY 1094 AAG-----CAGATCGAGCCCTCTCCCTTAATAATTA 1123
 Db 230 ArgGlyValGluAspValAspGlnGluIleGlnGlnIleLysAlaGlu---LeuGluAla 248
 QY 1124 ACATC-----ATGAGCTGCGTATTGTCGATGTTGTCAG 1162
 Db 249 ThrValSerGluGluLysAlaGlyLysAlaSerTyr-----GlyGluLeuPheArg 265
 QY 1163 AACAGGCTAGATTCTTACGGAACCTTAACGAACGAAATTTGGTCAAAGTGA 1222
 Db 266 GlyArgThrArg-----ProLysVal 272
 QY 1223 AGCAGCAACTATCGAAGAAATGCGTTATCAGCGACCATGCGGCTTTAGAGCTTGG 1282
 Db 273 ArgGlnArgLeuLeuMetGlyValMetLeuGlnAlaPheGlnGlnLeuThrGlyIleAsn 292
 QY 1283 GCATCATC-----TTGCTCTATGATGATGTCGCTTGAATGCAATATGCT 1330
 Db 293 AlaIlePheTyrTyrSerProThrIlePheLysSerValGlyValSerAspSerValAla 312
 QY 1331 -----TTCAGTCCGATGCGCTTAAATTCATGACCTTTGGCTACCTGTCAGCTC 1381
 Db 313 SerLeuLeuValThrIleIleValGlyValAlaAsnPheValPheThrPheValAlaLeu 332
 QY 1382 TTGTTTATAGCAGATTTCTTTTGAAGAAATTCAAATGATTTGCAAGCATTTGTCGT 1441
 Db 333 IlePheLeuValAspArgPheGlyArgArgProLeuLeuLeuLeuGlyAlaIleGlyMet 352
 QY 1442 TTAATG-----ACTGATTTGGGTTATTCATTAACAATTACTTGTGATTTTGAATCGT 1495
 Db 353 AlaIleCysPheLeuLeuLeuGlyAlaSerIleGlyValAlaLeuLeuLeuLeuAsnLys 372
 QY 1496 ATTGCTGA-----GATGCCAAGCGAACCTGTTTACCCCTATGCAATGTTTAT 1549
 Db 373 ProLysAspProSerSerLysAlaIleGlyIleValAlaIleValPheIleLeuLeuPhe 392
 QY 1550 GATGCC-----CTTCAAAAGAGCTTC 1570
 Db 393 IleAlaPhePheAlaLeuGlyTyrGlyProIleProTyrValIleLeuSerGlyLeuPhe 412
 QY 1571 AGCCGACGCGTAAATGACCAACAGCTACCAACTCTTCAAGTTTGTAAAGCTTTTG----- 1624
 Db 413 ProThrLysValArgSerLysAlaLeuAlaLeuAlaThrAlaIleAsnTyrPheAlaAsn 432

QY 1625 TTTATAGCGGCTCTCTGCTTTAAATTTTCATTTATTATG---ACCATAGGATCTTT 1681
 |||||
 Db 433 PheIleLeuGly-----PheLeuPheProTyrIleThrGlyAlaIleGlyLeuAla 449
 |||||
 QY 1682 CTAGAACTTTATCGTCTCTATATATTCACACCTCTGTTGTTGTTATGCTC 1735
 |||||
 Db 450 LeuGlyGlyTyrValPheLeuValPheAlaGlyLeuValLeuPheIleLeu 467

RESULT 29

US-10-095-139-14
 ; Sequence 14, Application US/10095139
 ; Patent No. US20020165357A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Curtis, Rory A.J.
 ; APPLICANT: Silos-Santiago, Immaculada
 ; APPLICANT: Millennium Pharmaceuticals, Inc.
 ; TITLE OF INVENTION: 38554, 57301, and 58324, Human Organic
 ; TITLE OF INVENTION: Ion Transporters and Uses Therefor
 ; FILE REFERENCE: MF101-017P1RNM
 ; CURRENT APPLICATION NUMBER: US/10/095,139
 ; CURRENT FILING DATE: 2002-03-11
 ; PRIOR APPLICATION NUMBER: 60/275,172
 ; PRIOR FILING DATE: 2001-03-12
 ; NUMBER OF SEQ ID NOS: 24
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 14:
 ; LENGTH: 487
 ; TYPE: PRT
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: consensus
 ; US-10-095-139-14

Alignment Scores:

Pred. No.: 0.0319 Length: 487
 Score: 102.50 Matches: 110
 Percent Similarity: 35.67% Conservative: 68
 Best Local Similarity: 22.04% Mismatches: 169
 Query Match: 3.09% Indels: 152
 DB: 9 Gaps: 20

US-09-868-987-1 (1-1864) x US-10-095-139-14 (1-487)

QY 596 TTGACTACGATTGGCTCAGCACTCTTTCTCTCCTAGATCAGGCGCTATTAAAGGG 655
 |||||
 Db 3 LeuValAlaLeuGlyGlyGlyPheLeuPheGlyTyrAspThrGlyValIleGly 22
 |||||
 QY 656 TTTCCTTTGCATTGATT----- 673
 |||||
 Db 23 Phe---LeuAlaLeuIleAspPheLeuPheArgPheGlyLeuLeuThrSerSerGlyAla 41
 |||||
 QY 674 -----TTAGGAATTTCTCTCAATGTTTACGGCTCTTTTCATGACTAAATTTTC 724
 |||||
 Db 42 LeuAlaGluLeuGly---TyrSerThrValLeuThrGlyLeuValSerIlePhePhe 60
 |||||
 QY 725 TTCATGCTGTGGATCAATPAGACCCACATACACAGTTGTCATGATGAATAAGTTCTGT 784
 |||||
 Db 61 LeuGlyArgLeuIleGly-----SerLeuPheAla 70
 |||||
 QY 785 GGGATAAGCATGATTCTTCAGAGGATGCAAAACCTTTGGCTGTTTCTGGAGTGTT 844
 |||||
 Db 71 GlyLysLeuGlyAspArgPhe---GlyArgLysLysSerLeuLeuIleAlaLeuValLeu 89
 |||||
 QY 845 TTCTTTTTCAGTTCGTCCTCGGTTTTCGAGCCTGGA-----ATTCCGTTTTC 895
 |||||
 Db 90 PheValIleGly-AlaLeuLeuSerGlyAlaAlaProGlyTyrThrThrIleGlyLeuTr 109
 |||||
 QY 896 GGAATGGATTTTA----- 908
 |||||
 Db 109 pAlaPheTyrLeuLeuIleValGlyArgValLeuValGlyLeuGlyValGlyAlaSe 129
 |||||
 QY 909 -----AAGGAGGTATGCTCTTT 925

Db 129 rValLeuValProMetTyrIleSerGluIleAlaProLysAlaLeuArgGlyAlaLeuGl 149
 |||||
 QY 926 ACCTTTAATCCAAAGACGATCGCATCGAGCTGTTGCTCCTCAAAATCGCTGCAAAAGTTGTG 985
 |||||
 Db 149 ySerLeuTyrGlnLeuAlaIleThrIleGlyIleLeuValAlaAlaIleIleGlyLeuGl 169
 |||||
 QY 986 CATAACTACAGGAG----- 1001
 |||||
 Db 169 yLeuAsnLysThrAsnAsnAspSerAlaLeuAsnSerTrpGlyTrpArgIleProLeuGl 189
 |||||
 QY 1002 -----CTGGTCTTTCTCTAGAGACTTCCTCGATTCAAA 1033
 |||||
 Db 189 yLeuGlnLeuValProAlaLeuLeuLeuIleGlyLeuLeu-PheLeuProGluSerP 209
 |||||
 QY 1034 ACATTTGGATCTTCAGAAAAGATCAAAATCTATTATTTAGTGATAAAGCTTTTAAGCTATACT 1093
 |||||
 Db 209 roArgTrpLeuVal-GluLysGlyLysLeuGluAlaArgGluValLeuAlaLysLeu 228
 |||||
 QY 1094 AAG-----CAGATACGAGCCTCTCTC----- 1114
 |||||
 Db 229 ArgGlyValGluAspValAspGlnGluIleGlnGluIleLysAlaGluLeuGluAlaGly 248
 |||||
 QY 1115 -----CTAAATTAACGATCATGCTGCGGTTATTGCTGGGATTGTTGTGAGAAC 1165
 |||||
 Db 249 ValGluGluGluLysAlaGlyLysAlaSerTrp-----GlyGluLeuPheArgGly 265
 |||||
 QY 1166 AGGCTAGATTTCTCTACGGAAACTCTAAACGAAACGAAATTTTGGTCAAGGTAAAGC 1225
 |||||
 Db 266 ArgThrArg-----ProLysValArg 272
 |||||
 QY 1226 AGCAACTATCGAAGAAATGCGTTATCAGCGCAACATCGGGCTTTTAGAGCTTTGGCA 1285
 |||||
 Db 273 GlnArgLeuLeuMetGlyValMetLeuGlnAlaPheGlnGlnLeuThrGlyIleAsnAla 292
 |||||
 QY 1286 ATCATC-----TTGCTCTATGTGAGTTTGGCTTTTGAATGCAATATGCT--- 1330
 |||||
 Db 293 IlePheTyrTyrSerProThrIlePheLysSerValGlyValSerAspSerArgAlaSer 312
 |||||
 QY 1331 -----TTCAGTCGCTGATGCGCTTTAATTCATGACCTTTTGGCTACCTGCGCAGTCTTG 1384
 |||||
 Db 313 LeuLeuValThrIleIleValGlyValValAsnPheValPheThrLeuValAlaLeuIle 332
 |||||
 QY 1385 TTTATACACATTTCTTTTGAAGAAATTCAAATAGATTTCAGGCAATTTGGTCTTTA 1444
 |||||
 Db 333 PheLeuValAspArgPheGlyArgArgProLeuLeuLeuGlyAlaAlaGlyMetAla 352
 |||||
 QY 1445 ATG-----ACTGTATTGGGTATTTCATTAAACAATCTTTGATCATTTTTCGCTATT 1498
 |||||
 Db 353 IleCysPheLeuIleLeuGlyAlaSerIleGlyValAlaLeuLeuLeuLeuAsnLysPro 372
 |||||
 QY 1499 CGTGAA-----GATCGCAAGCGAACTGTTTACCCTCATGCTTTTGTAGTTAATCAT 1552
 |||||
 Db 373 LysAspProLeuSerLysAlaAlaGlyIleValAlaIleValPheIleLeuLeuPheIle 392
 |||||
 QY 1553 GCC-----CTTCAAAAGACGTTTCAGC 1573
 |||||
 Db 393 AlaPhePheAlaLeuGlyTrpGlyProIleProTrpValIleLeuSerGluLeuPhePro 412
 |||||
 QY 1574 CGCAGGTAAATGACACAGCTACAACTCATCTAGTTTGTGTTTGTGCTTTTG-----TTT 1627
 |||||
 Db 413 ThrLysValArgSerLysAlaLeuAlaLeuAlaThrAlaAlaAsnTrpLeuAlaAsnPhe 432
 |||||
 QY 1628 ATAGCGGCTCTCTGCTTTTAAATTTTGCATTTTATTATG---ACCATAGGATCTCTCTA 1684
 |||||
 Db 433 IleIleGly-----PheLeuPheProTyrIleThrGlyAlaIleGlyLeuAlaLeu 449
 |||||
 QY 1685 GGAATCTTATGCTCTCTTTATATTGACCACTCTGTTGTTGTTTATGGTC 1735
 |||||
 Db 450 GlyGlyTyrValPheLeuValPheAlaGlyLeuLeuValLeuPheIleLeu 466
 |||||

RESULT 30

US-09-795-693-27

```

; Sequence 27, Application US/09795693
; Patent No. US20020068710A1
; GENERAL INFORMATION:
; APPLICANT: Glucksmann, Maria A.
; TITLE OF INVENTION: 20685, 579, 17114, 23821, 33894, and
; TITLE OF INVENTION: 32613, No. US20020068710A1 Human Transporters
; FILE REFERENCE: 35800/209292
; CURRENT APPLICATION NUMBER: US/09/795,693
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: 60/185,906
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 27
; LENGTH: 487
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Pfam consensus sequence
; US-09-795-693-27

Alignment Scores:
Pred. No.: 0.0319 Length: 487
Score: 102.50 Matches: 110
Percent Similarity: 35.67% Conservative: 68
Best Local Similarity: 22.04% Mismatches: 169
Query Match: 3.05% Indels: 152
Gaps: 20

US-09-868-987-1 (1-1864) x US-09-795-693-27 (1-487)

QY 596 TTGACTACAGTATGGCTTCAGACTTTCTTTCTTCTAGATACAGGCGCTTTAAAGG 655
Db 3 LeuValAlaAlaLeuGlyGlyPheLeuPheGlyTyAspThrGlyValIleGlyGly 22
QY 656 TTTCGTTGACATGATTT----- 673
Db 23 Phe---LeuAlaLeuIleAspPheLeuPheArgPheGlyLeuLeuThrSerSerGlyAla 41
QY 674 -----TTAGCAATTTCTCTCTCATGTTTACGGCTCTTTTCATCAATAATTTTTC 724
Db 42 LeuAlaGluLeuGly---TyrSerThrValLeuThrGlyLeuValValSerIlePhePhe 60
QY 725 TTCATGCTGTGGATGAATTAAGCCACACATACACAGTTGCATGATGATGAATGTTGCTG 784
Db 61 LeuGlyArgLeuIleGly-----SerLeuPheAla 70
QY 785 GGGATTAAGCATGATTTCTTGAGAGATGCAAAAACCTTGCGCTTTTCGAAGCTT 844
Db 71 GlyIysLeuGlyAspArgPhe---GlyArgIysIysSerLeuLeuIleAlaLeuValLeu 89
QY 845 TTTCCTTAGGTTGGCTGCTCTCGGTTTGGAGCCTGA-----ATTCCGTTTTCG 895
Db 90 PheValIleGlyAlaLeuLeuSerGlyAlaAlaProGlyIyrThrThrIleGlyLeuTr 109
QY 896 GGAATGATTTTA----- 908
Db 109 PALAPheTyLeuLeuIleValIleGlyArgValLeuValGlyLeuGlyValGlyAlaLase 129
QY 909 -----AAGAGGGGATGCTTT 925
Db 129 rValLeuValProMetTyIleSerGluIleAlaProIysAlaLeuArgGlyAlaLeuGly 149
QY 926 ACCTTAATCCAAAGACATGCGATGCGATGCGATGCGTCAATGCGTGCAGAAAGTTGTG 985
Db 149 ySerLeuTyGlnLeuAlaIleThrIleGlyIleLeuValAlaAlaIleIleGlyLeuGly 169
QY 986 CATAAACTACAGAGAG----- 1001
Db 169 yLeuAsnIysThrAsnLeuAspSerAlaLeuAsnSerTrpGlyTrpArgIleProLeuGly 189
QY 1002 -----CTGCTCTTTCTTCTAGAGACTTCGCTATTGCA 1033

```

```

Db 189 yLeuGlnLeuValProAlaLeuLeuLeuIleGlyLeuLeu-PheLeuProGlyLaseP 209
QY 1034 ACATTTGGATCTTCAGAAAAGATCAAAATCTATTATGTCATTAAGCTTTAAGCTTACT 1093
Db 209 roArgTrpLeuVal-GlyIysGlyIysLeuGlnGluAlaArgGluValAlaIleAlaIysLeu 228
QY 1094 AAG-----CAGTACAGCCCTCTC----- 1114
Db 229 ArgGlyValGluAspValAspGlnGluIleGlnGluIleIysAlaGluLeuGluAlaGly 248
QY 1115 -----CTAAATTAAGATCATGAGCTGCGTATGTTGGATGTTGTCAGAAAC 1165
Db 249 ValGluGluGluIysAlaGlyIysAlaSerTrp-----GlyGluLeuPheArgGly 265
QY 1166 AGGCTTAGATTTCTTAGGAAACTTAACGAAACGAAAAATTTGCTCAAGGTATAC 1225
Db 266 ArgThrArg-----ProIysValArg 272
QY 1226 AGCAACTATCGAAGAAATGCGTTATCAGCGACCATGCGGCTTTTGAAGCTTTGCA 1285
Db 273 GlnArgLeuLeuMetGlyValMetLeuGlnAlaPheGlnGlnLeuThrGlyIleAsnAla 292
QY 1286 ATCATC-----TTGCTCTATGATGATGTTGGCTTTGAATGCAATATGCT--- 1330
Db 293 IlePheTyTrpSerProThrIlePheIlePheIleSerValGlyValSerAspSerArgAlaSer 312
QY 1331 -----TTCAAGTCCGATGCGCTTATATTCATGACCTTTGGCTACCTGTCAGTCTTG 1384
Db 313 LeuLeuValThrIleIleValIleGlyValAlaAsnPheValPheThrLeuValAlaLeuIle 332
QY 1385 TTATATACACATTTCTTTTGAAGAAATCAATATGATTTGCAAGCATTTGGCTTTA 1444
Db 333 PheLeuValAspArgPheGlyArgArgProLeuLeuLeuLeuGlyAlaAlaGlyMetAla 352
QY 1445 ATG-----ACTGATTTGGGATTTTCATTAACAATCTTTGATCATTTTTCATGCTAT 1498
Db 353 IleCysPheLeuIleLeuGlyAlaSerIleGlyValAlaLeuLeuLeuLeuAsnIysPro 372
QY 1499 CGTGA-----GATGCCACAGGAGACCTGTTTACCCCTATGATGATTTTATGAT 1552
Db 373 LysAspProLeuSerIysAlaAlaIleValAlaIleValPheIleLeuPheIle 392
QY 1553 GCC-----CTTCAAAAGACGTTGACG 1573
Db 393 AlaPhePheAlaLeuGlyIyrGlyProIleProIyrValIleLeuSerGluLeuPhePro 412
QY 1574 CGCAGCGTAAATGACAAACAGCTACACACTATCACTTTTGTAAATGCTTTG-----TTT 1627
Db 413 ThrIysValArgSerIysAlaLeuAlaLeuAlaThrAlaAlaAsnTrpLeuAlaAsnPhe 432
QY 1628 ATAGCGCGCTCTCTGCTTTTAATTTTGCATTTATATG--ACCATAGGATTTCTTCTA 1684
Db 433 IleIleGly-----PheLeuPheProTyIleThrGlyAlaIleGlyLeuAlaLeu 449
QY 1685 GGAATTTATTCGCTCTTTATATTTGACACACCTGTTGTTGTTTATGATGTC 1735
Db 450 GlyIlyTyValPheLeuValPheAlaGlyLeuLeuValLeuPheIleLeu 466

RESULT 31
US-09-827-937A-18
; Sequence 18, Application US/09827937A
; Patent No. US20020052043A1
; GENERAL INFORMATION:
; APPLICANT: Li, Yi
; APPLICANT: Ruben, Steven M.
; TITLE OF INVENTION: Human G-Protein Coupled Receptors
; FILE REFERENCE: 1488.1220003
; CURRENT APPLICATION NUMBER: US/09/827,937A
; PRIOR FILING DATE: 2001-04-09
; PRIOR APPLICATION NUMBER: 08/852,824
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Patent In Ver. 2.0

```



```
; SEQ ID NO 18
; LENGTH: 381
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-827-937A-18

Alignment Scores:
Pred. No.: 0.0409 Length: 381
Score: 101.00 Matches: 83
Percent Similarity: 31.83% Conservative: 58
Best Local Similarity: 18.74% Mismatches: 116
Query Match: 3.05% Indels: 186
DB: 22 Gaps: 22

US-09-868-987-1 (1-1864) x US-09-827-937A-18 (1-381)

QY 290 ATCTCAGCATGCTGCTGGCAATGCTTATTTGATGAGCGTATATTATAGATTT 349
Db 53 lLeuLeuCysCys-----PheIleLeuGluAsnIlePheVal----- 66

QY 350 GGAGCGCTCATCGCTTCGGAGCTGTTCTTCTGAATCTTTTGTCTATCTGGCAGCT--- 406
Db 67 -----LeuLeuThrIleTrpLysThrLys 74

QY 407 -----CTACAGTATTGGATGCGCCACTCACCTTGTGCAAGCTCGCTGGG 451
Db 75 LysPheHisArgProMetTyrPheIleGlyAsnLeuAlaLeuSerAspLeuLeuAla 94

QY 452 ATTGTTCTTGTATGGGATGCGGTAGATGCAAAATGTTCTTGTATTGCAAGAAATCCGA 511
Db 95 -----GlyValAlaTyrThrAlaAsnLeuLeu----- 104

QY 512 GAGGAATTTTATTGTCTCAAAAGTCTTAAATAATCTGTAGAAAAAGGATATACCAAGCT 571
Db 104 ----- 104

QY 572 TTGGAGCCATTTTGTATCTTAACCTGACTACATATGGCTCAGCACTTCTTTTCTTC 631
Db 105 -----SerGlyAlaThrThrTyrLysLeuThrProAlaGlnTrpPhe 118

QY 632 CTAGATACAGGCGCTATTAAAGGTTTTCCTTGACATTGATTTTGGAAATTTTCTCTTCA 691
Db 119 LeuArgGluGly-----SerMetPheValAlaLeuSerAlaSer 131

QY 692 ATGTTTACG-----GCTCTTTTCATGACTAAATTTTCTTCATGCTGTGGATG--- 739
Db 132 ValPheSerLeuLeuAlaIleAlaIleGluArgTyrIleThrMetLeuLysMetLysLeu 151

QY 740 ---AATAAGCCCAACATACACAGTTGCTATATGATGAATAAGTTCTGGGGATAAAGCAT 796
Db 152 HisAsnGlySerAsnAsnPheArgLeuPheLeu----- 163

QY 797 GATTCTTGTAGAGGATGCAAAACCTTTGGCTGTTTCTGGAAGTGTTTTCTTTTAGGT 856
Db 164 -----IleSerAlaCys-----TrpValIleSer-----LeuIleLeuGly 175

QY 857 TGCCTGTCTCGGCTTGGAGCTTGGAAATTCCTGTTTGGGAATCGATTTTAAAGGAGGG 916
Db 176 -----GlyLeuProIleMetGlyTrpAsnCysIle----- 185

QY 917 TATGCTTTTACCTTTAATCCAAAAGAGCATGGCATACGCGATGTTGTCAAATGCGTGGC 976
Db 185 ----- 185

QY 977 AAAGTTGTCATAAATACTACAGGAAGCTGGTCTTCTTCTAGAGACTTCCTGTAACAACA 1036
Db 186 -----SerAlaLeuSerSer----- 190

QY 1037 TTTGGATCTTCAGAAAAGATCAAAATCTATTATTAGTGATAAAGCTTTAAGCTATACTAAG 1096
Db 191 -----CysSerThrValLeuProLeuTyrHisLysHisTyrIleLeuPheCysThrThr 208

QY 1097 CAGATACAGGCTCTCTCCTTAAATAACGATCATGAGCTGGCGCTTATTGTGGGATT--- 1153
```

```
Db 209 ValPheThrLeuLeuLeuSerIleValIleLeu-----TyrCysArgIleTyr 225
QY 1154 ---GTTGTGAGAAAACAGGCTAGATTTCTCTACGGAAACTCTTAAACGAAACGCAAAATTT 1210
Db 226 SerLeuValArgThrArgSerArgArgLeu-----ThrPheArgLysAsn----- 240
QY 1211 TGTCAAAGGTAAAGCAGCAAACTATCGAAGAAAATCGTTTATCAGGCGACCATCGGCTT 1270
Db 241 -----IleSerLysAlaSerArgSerGluAsnValAlaLeu 253
QY 1271 TTAGGAGCTTTGGCAATCATCTTG---CCTATGTGAGTTTGGCTTTGAATGCAATAT 1327
Db 254 LeuLysThrValIleIleValLeuSerValPheIleAlaCys-----TrpAlaPro 270
QY 1328 GCTTTAGTCCCTATGCGCTTTAAATTTCATGACCTTTTGGCTACCTGTGCGACTTCTTT 1387
Db 271 LeuPheIleLeuLeuLeuLeuAspValGlyCysLysValLysThrCysAspIleLeuPhe 290
QY 1388 ATAGCACATTTCTTTTGAAGAAAATTCAAATAGATTTCGAAGCCATTGGTGTCTTAAATG 1447
Db 291 ArgAlaGluTyrPheLeu-----ValLeu 298
QY 1448 ACTGTATTGGGTATTTCATTAAACAATACTTTGATCATTTTTCATCGTATTCGTGAAGAT 1507
Db 299 AlaValLeuAsnSerGlyThrAsnProIleIle----- 309
QY 1508 CGCAAGCGAAACCTGTTTACCCCTATGCTGCTTTAGTTTAAATGATGCCCTTCAAAAGACG 1567
Db 310 -----TyrThrLeuThrAsnLysGluMetArgArgAla 320

QY 1568 TTCAGCGCG 1576
Db 321 PheIleArg 323

RESULT 32
US-09-842-316-6
; Sequence 6, Application US/09842316
; Patent No. US20020099191A1
; GENERAL INFORMATION:
; APPLICANT: KOSTENSIS, Eva
; APPLICANT: GASSENHUBER, Johann
; TITLE OF INVENTION: ED88 RECEPTOR, ITS PREPARATION AND USE
; FILE REFERENCE: 38005-147
; CURRENT APPLICATION NUMBER: US/09/842,316
; CURRENT FILING DATE: 2001-04-26
; PRIOR APPLICATION NUMBER: EP 116589.3
; PRIOR FILING DATE: 2000-08-01
; PRIOR APPLICATION NUMBER: EP 108858.2
; PRIOR FILING DATE: 2000-04-26
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
; LENGTH: 381
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-842-316-6

Alignment Scores:
Pred. No.: 0.0409 Length: 381
Score: 101.00 Matches: 83
Percent Similarity: 31.83% Conservative: 58
Best Local Similarity: 18.74% Mismatches: 116
Query Match: 3.05% Indels: 186
DB: 22 Gaps: 22

US-09-868-987-1 (1-1864) x US-09-842-316-6 (1-381)

QY 290 ATCTCAGCATGCTGCTGGCAATGCTTATTTGATGAGCGTATATTATAGATTT 349
Db 53 lLeuLeuCysCys-----PheIleLeuGluAsnIlePheVal----- 66

QY 350 GGAGCGCTCATCGCTTCGGAGCTGTTCTTCTGAATCTTTTGTCTATCTGGCAGCT--- 406
```

```

Db 67 -----LeuLeuThrIleTyrPheLeu 74
Qy 407 -----CTACAGTATTGGATGGCCATCACCCTTGTCAGACTCGCTGG 451
Db 75 LysPheHisArgProMetTyrTyrPheIleGlyAsnLeuAlaLeuSerAspLeuAla 94
Qy 452 ATTGTTCTTGCTATGGGATGGCCATGCAATGTTCTTGATTCGAAAGAAATCCGA 511
Db 95 -----GlyValAlaTyrThrAlaAsnLeuLeu----- 104
Qy 512 GAGGAATTTTATTGTCTCAAAAGTCTTAAAAATCTGAGAAAAAGATATACCAAGGCT 571
Db 104 ----- 104
Qy 572 TTGGAGCCATTTTGTATCTAATCTGACCTAGATGGCCCTGACACTCTTTCTTC 631
Db 105 -----SerGlyAlaThrThrTyrLysLeuThrProAlaGlnTyrPhe 118
Qy 632 CTAGATACAGGCGCTATTAAAGGTTTGTGACATTGATTTTGAATTTTCTCTTCA 691
Db 119 LeuArgGly-----SerMetPheValAlaLeuSerAlaSer 131
Qy 692 ATGTTTACG-----GCTCTTTTCATGACTAATTTTCTTCATGCTGTGATG----- 739
Db 132 ValPheSerLeuLeuAlaIleAlaIleGluArgTyrIleThrMetLeuLysMetLysLeu 151
Qy 740 ---AATAGACCCACATACACAGTTGATGATGAAATTAAGTTGCTGGGATAAAGCAT 796
Db 152 HisAsnGlySerAsnAsnPheArgLeuLeuLeu----- 163
Qy 797 GATTCTTGAAGAGATGCAAAAAAATTGGGCTGTTTCTGCAAGCTTTTCTTTAGT 856
Db 164 -----IleSerAlaCys-----TyrValIleSer-----LeuIleLeuGly 175
Qy 857 TGGCTTGCTCTCGGTTTGGAGCTGGAATTCGGTTTGGGATGATTTTAAAGAGAGG 916
Db 176 -----GlyLeuProIleMetGlyTyrAsnCysIle----- 185
Qy 917 TATGCTTTACCTTAAATCAAAAGAGATGACATGACATGTTGCTCAATGCGTGGC 976
Db 185 ----- 185
Qy 977 AAAGTTTGCACTAACTACAGAACTGCTTTTCTTACAGACTCCGATTTCAAACA 1036
Db 186 -----SerAlaLeuSerSer----- 190
Qy 1037 TTGGATCTTCAGAAAAATCAAAATCTAATTTAGTAAAGCTTTAAGCTATCTAAG 1096
Db 191 -----CysSerThrValLeuProLeuTyrHisLysHisTyrIleLeuPheCysThrThr 208
Qy 1097 CAGATACAGAGCTCTCTCTAATAATTAACATGATGACGCTGGCTTATGTGGGAT-- 1153
Db 209 ValPheThrLeuLeuLeuSerIleValIleLeu-----TyrCysArgLysIleTyr 225
Qy 1154 ---GTTGTCAGAAACAGCCTTAGATTTCTTACGGAACCTTAAACGAAACCAAAATTT 1210
Db 226 SerLeuValArgThrArgSerArgLeu-----ThrPheArgLysAsn----- 240
Qy 1211 TGTCAAAAGTAAAGAGCAAACTATCGAAGAAATGCTTACGAGCAGCAACGCGGCTT 1270
Db 241 -----IleSerLysAlaSerArgSerSerGluAsnValAlaLeu 253
Qy 1271 TTAGAGCTTTGGCAATCATCTTG---CTCATGTGAGTTTGGCTTGAATGGCAATAT 1327
Db 254 LeuLysThrValIleIleValLeuSerValPheIleAlaCys-----TyrAlaPro 270
Qy 1328 GCTTCACTGCCGATAGCGCTTAAATCATGACCTTTGGCTACTGTGCACTCTGTTT 1387
Db 271 LeuPheIleLeuLeuLeuAspValGlyCysLysValLysThrCysAspIleLeuPhe 290
Qy 1388 ATAGACATTTCTTTTGAAGAAATTCAAATAGATTTTGAAGCAATGGGCTTATATG 1447

```

```

Db 291 ArgAlaGlyTyrPheLeu-----ValLeu 298
Qy 1448 ACTGATTTGGGGATTTATTAAACATCTTGTATCATTTTGTATGCTATTCTGAAGAT 1507
Db 299 AlaValLeuAsnSerGlyThrAsnProIleIle----- 309
Qy 1508 CGCCACGCAACCGTTTACCCCTATGCTTTTATGATGATGCGCTTCAAAAGACG 1567
Db 310 -----TyrThrLeuThrAsnLysGluMetArgAla 320
Qy 1568 TTCAGCCGC 1576
Db 321 PheIleArg 323

RESULT 33
US-09-731-030A-17
; Sequence 17, Application US/09731030A
; Patent No. US20020142375A1
; GENERAL INFORMATION:
; APPLICANT: MURROE, Donald G
; APPLICANT: GUPTA, Ashwani K.
; APPLICANT: ZASTAMNY, Roman L.
; TITLE OF INVENTION: MAMMALIAN EDG-7 RECEPTOR HOMOLOGS
; FILE REFERENCE: 8074-0015
; CURRENT FILING DATE: 1998-12-29
; PRIOR APPLICATION NUMBER: 60/070,184
; PRIOR FILING DATE: 1997-12-30
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 17
; LENGTH: 381
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: Mammalian EDG
US-09-731-030A-17

Alignment Scores:
Pred. No.: 0.0409 Length: 381
Score: 101.00 Matches: 83
Percent Similarity: 31.83% Conservative: 58
Best Local Similarity: 18.74% Mismatches: 116
Query Match: 3.05% Indels: 186
DB: Gaps: 22

US-09-868-987-1 (1-1864) x US-09-731-030A-17 (1-381)
Qy 290 ATCTCACATGCTGTGCTGGCAATCTTATTGTTTGAAGCGATATTATATGATTT 349
Db 53 IleLeuIleCysCys-----PheIleIleLeuGluAsnIlePheVal----- 66
Qy 350 GAGGCGCTCATCGCTTGGAGAGCTGTTCTGATATTTTGTGTTATCTGGGAGCT-- 406
Db 67 -----LeuLeuThrIleTyrPheLys 74
Qy 407 -----CTACAGTATTGGATGGCCACTCACTTGTGACAGACTCGCTGG 451
Db 75 LysPheHisArgProMetTyrTyrPheIleGlyAsnLeuAlaLeuSerAspLeuAla 94
Qy 452 ATTGTTCTTGCTATGGGATGGCCATGCAATGTTCTTGATTCGAAAGAAATCCGA 511
Db 95 -----GlyValAlaTyrThrAlaAsnLeuLeu----- 104
Qy 512 GAGGAATTTTATTGTCTCAAAAGTCTTAAAAATCTGAGAAAAAGATATACCAAGGCT 571
Db 104 ----- 104
Qy 572 TTGGAGCCATTTTGTATCTAATCTGACCTAGATGGCCCTGACACTCTTTCTTC 631
Db 105 -----SerGlyAlaThrThrTyrLysLeuThrProAlaGlnTyrPhe 118
Qy 632 CTAGATACAGGCGCTATTAAAGGTTTGTGACATTGATTTTGAATTTTCTCTTCA 691

```


; APPLICANT: Chu, Peter
 ; APPLICANT: Pardo, Jorge
 ; APPLICANT: Li, Congfen
 ; APPLICANT: Zhao, Haoran
 ; APPLICANT: Jiang, Yingping
 ; APPLICANT: Rigel Pharmaceuticals, Incorporated
 ; TITLE OF INVENTION: EDG: Modulators of Lymphocyte Activation and Migration
 ; FILE REFERENCE: 021044-000310US
 ; CURRENT APPLICATION NUMBER: US/09/971,228
 ; CURRENT FILING DATE: 2002-04-19
 ; PRIOR APPLICATION NUMBER: US 60/284,763
 ; PRIOR FILING DATE: 2001-04-18
 ; NUMBER OF SEQ ID NOS: 15
 ; SOFTWARE: PatentIn ver. 2.1
 ; SEQ ID NO 13
 ; LENGTH: 341
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; OTHER INFORMATION: mutant #1, C-terminally truncated variant of
 ; OTHER INFORMATION: human endothelial differentiation G-protein
 ; OTHER INFORMATION: coupled receptor (GPCR) 1 (EDG1)
 ; US-09-971-228-13

Alignment Scores:

Pred. No.: 0.112 Length: 341
 Score: 96.50 Matches: 83
 Percent Similarity: 32.05% Conservative: 59
 Best Local Similarity: 18.74% Mismatches: 116
 Query Match: 2.91% Indels: 185
 DB: 9 Gaps: 22

US-09-868-987-1 (1-1864) x US-09-971-228-13 (1-341)

QY 290 ATCTCAGCATGCTGGGCTTGGCAATGCTTATGTTTGGATGAGCGTATATATAGATT 349
 Db 53 lileulleCysCys-----PheilleleuGluasnillepheval----- 66
 QY 350 GGAGCGTCATCGCTTCGGGAGCTGTTCTTCTGAATCTTTTCTTATCTGGGCAGCT--- 406
 Db 67 -----LeuLeuThrlleTlpylsThrllys 74
 QY 407 -----CTACAGTATTGGATGGCCACTCACCCTGTCTAGGACTCGCTGGG 451
 Db 75 LysPheHisArgProMetTyrTyrPheilleGlyAenLeuAlaLeuSerAspLeuLeuAla 94
 QY 452 ATTGTTCTTGCTATGGGATGCGCTAGATCAATGTTCTGTATTTCGAAGATCCGA 511
 Db 95 -----GlyValAlaTyrThrAlaAenLeuLeu----- 104
 QY 512 GAGGAATTTTATTGCTCAAAAGTCTTAAATAATCTGTAGAAAAAGGATATACCAAGGCT 571
 Db 104 ----- 104
 QY 572 TTGGAGCATTTTGGATCTAACTGATACAGTATTGGCTCAGCACTTCTTTTCTTC 631
 Db 105 -----SerGlyAlaThrTyrLysLeuThrProAlaGlnTrpPhe 118
 QY 632 CTAGATACAGGCCTATTAAAGGTTTCTTGCATGATTTAGGAATTTCTCTTCA 691
 Db 119 LeuArgGluGly-----SerMetPheValAlaLeuSerAlaSer 131
 QY 692 ATGTTTACG-----GCTCTTTTCATGACTAAATTTTCTTCATGCTGTGGATG----- 739
 Db 132 ValPheSerLeuLeuAlaAlaileGluArgTyrilleThrMcLeuLysMetLysLeu 151
 QY 740 ---AATAGACCCAAATACACAGTTGCATATGATGAATAAGTTCGTGGGATAAGCAT 796
 Db 152 HisAenGlySerAsnAenPheArgLeuPheLeuLeu----- 163
 QY 797 GATTCTCTGAGGAGTGAATAAACTTTGGGCTGTTCTGGAAGTGTCTTTCTTTAGGT 856
 Db 164 -----ileSerAlaCys-----TrpValileSer-----LeuileuGly 175

QY 857 TGCCTTCTCTCGGGTTTGGAGCCTGGAATTCGCTTTTGGGAATGGATTTTAAAGAGGG 916
 Db 176 -----GlyLeuProilleMetGlyTrpAsnCysile----- 185
 QY 917 TATGCTTTTACCTTTAATCCAAAGAGCATGGCATCAGCGATGTGTCTCAATCGGTGGC 976
 Db 185 ----- 185
 QY 977 AAAGTTGTGCATAAACTACAGGAAGCTGGTCTTCTTTCTAGAGACTTCCGCTATTCAAACA 1036
 Db 186 -----SerAlaLeuSerSer----- 190
 QY 1037 TTGGATCTTCAGAAAAGATCAAAATCTATTTTAGTGATAAAGCTTTTAAGCTATATAAG 1096
 Db 191 -----CysSerThrValLeuProLeuTyrHisLysHisTyrilleLeuPheCysThrThr 208
 QY 1097 CAGATACAGCCTCTCTCTAAATAATTAACCATCATGAGCTGGCGTTATTGGGATT--- 1153
 Db 209 ValPheThrLeuLeuLeuSerilleValileLeu-----TyrCysArgileTyr 225
 QY 1154 ---GTTGTCAGAAACAGCCTAGATTTCTCTACGGAACCTCTAAACGAAACGCAAAATTT 1210
 Db 226 SerLeuValArgThrArgSerArgArgLeu-----ThrPheArgLysAenilleSerLys 243
 QY 1211 TGGTCAAAAGGTAAAGCAGCAAACTATCGAAGAAAATGCGTTTATCAGGCGACCATCGGCTT 1270
 Db 244 AlaSerArgSerSerGluLys-----SerLeuAlaLeu 254
 QY 1271 TTAGGAGCTTTGGCAATCATCTTG---CTCATGTGAGTTTGGCCTTTGAATGCAATAT 1327
 Db 255 LeuLysThrValilleValLeuSerValPheillealacys-----TrpAlaPro 271
 QY 1328 GCTTTCAGTGCGGTATCGCTTTTAAATTCATGACCTTTTGGTACCTGTGCACTCTTGT 1387
 Db 272 LeuPheilleLeuLeuLeuLeuAepValGlyCysLysValLysThrCysAspilleLeuPhe 291
 QY 1388 ATAGCACATTTCTTTTGAAGAAAATTCAAATAGATTTGCAAGCCATTGGTGTCTTAAATG 1447
 Db 292 ArgAlaGluTyrPheLeu-----ValLeu 299
 QY 1448 ACTGATTGGGTATTTCATTAACAATACTTTGATCATTTTGTATCGTATTTCGTGAAGAT 1507
 Db 300 AlaValLeuAenSerGlyThrAsnProille----- 310
 QY 1508 CGCCAAGCAACCTGTTTACCCTTATGCATGTTTGTAGTAAATGATGCCCTTCAAAAGACG 1567
 Db 311 -----TyrThrLeuThrAenLysGluMetArgGala 321
 QY 1568 TTCAGCGCGC 1576
 Db 322 PheileArg 324

RESULT 39

; US-09-971-228-15
 ; Sequence 15, Application US/09971228
 ; Patent No. US20020155512A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Liao, X. Charlene
 ; APPLICANT: Masuda, Esteban
 ; APPLICANT: Chu, Peter
 ; APPLICANT: Pardo, Jorge
 ; APPLICANT: Li, Congfen
 ; APPLICANT: Zhao, Haoran
 ; APPLICANT: Jiang, Yingping
 ; APPLICANT: Rigel Pharmaceuticals, Incorporated
 ; TITLE OF INVENTION: BDG: Modulators of Lymphocyte Activation and Migration
 ; FILE REFERENCE: 021044-000310US
 ; CURRENT APPLICATION NUMBER: US/09/971,228
 ; CURRENT FILING DATE: 2002-04-19
 ; PRIOR APPLICATION NUMBER: US 60/284,763
 ; PRIOR FILING DATE: 2001-04-18
 ; NUMBER OF SEQ ID NOS: 15

QY	290	ATTCAGACAGCTCGGGCTTGGCAATCCCTAATGTTTGTATGATGGCCTAATATATGATT	349
Db	53	IleuIleIecysCys-----PheIleIeuGlubasnIleheval-----	66
QY	350	GGAGCGCTCATCGCTTCGGAGAGCTGTTCTTGTAACTCTTTGCTATTCGGGCACT--	406
Db	67	-----LeuIeuThrIleThrIpySthIrys	74
QY	407	-----CTACGATATTGGATGGCCCACTACCTTGTCCAGACTTCGCTGGG	451
Db	75	LysPheHisargPrometIYrTYrPheIleGlyAsnIeuAlaIeuSerAspIeuIeuAla	94
QY	452	ATTGTTCTTGCTATGATGGAGCGCGGTAGATGCAAAATGTTCTGTATGCAAAAGATCCGA	511
Db	95	-----GlyValAlaIYrThrAlaAsnIeuIeu-----	104
QY	512	GAGGAAATTTTATTGTCTCAAAAGCTTTAAAAATCTGTAGAAAAAGATATACCAAGCT	571
Db	104	-----	104
QY	572	TTTGAGGCCATTTTGAATCTTAACTTACTACAGATATGGCTCCAGACTCTTTCTTC	631
Db	105	-----SergIyAlaThrThrTYrIlyIysIeuThrProIaGlnIrpPhe	118
QY	632	CTAGATACAGGCGCTATTAAAGGTTTGCTTGCATGATTTAGAAATTTCTCTCA	691
Db	119	LeuargIgluIy-----SerMetheValAlaIeuSerAlaSer	131
QY	692	ATGTTTAGC-----GCTCTTTCATGACTAAATTTTCTTCATGCTGTGATG-----	739
Db	132	ValPheSerIeuLeuAlaIleAlaIleGluArgTYrIleTherMetIeuIySmetIyIeu	151
QY	740	---AATTAAGCCCAACATACACAGTTCGATATGATTAAGTTCGTGGGATAAAGCAT	796
Db	152	HisAsnGlySerAsnAsnPheArgIeuPheIeuIeu-----	163
QY	797	GATTTCTTGAGAGGATGCAAAAACTTTGGCGCTGTTCTCGAAGCTGTTTTCTTTAGGT	856
Db	164	-----IleSerAlaCys-----TrpValIleSer-----LeuIleIeuGly	175
QY	857	TGCGTTCCTCTCGGGTTTGAGACCTCGGAATTCGGTTTGGAAATGATTTTAAAGAGGG	916
Db	176	-----GlyIeuProIleMetGlyTrpAsnCySile-----	185
QY	917	TATGCTTTACCTTTAATCCAAAAAGACATGACATACGATGTTGCTCAAAATGCGTGC	976
Db	185	-----	185
QY	977	AAAGTTTGCAATAACTACAGAGAGCTGGCTTTCTTCTAGAGACTTCGGTATTCACACA	1036
Db	186	-----SerAlaIeuSerIeu-----	190

Best Local Simi

Pred. No.:	0.115	Length:	362
Score:	96.50	Matches:	83
Percent Similarity:	32.05%	Conservative:	59
Best Local Similarity:	18.74%	Mismatches:	116

Query Match: 2.91% Indels: 185
DB: 10 Gaps: 22

US-09-868-987-1 (1-1864) x US-09-993-844-8 (1-362)

QY 290 ATCTCAGCATCTGGCTTGGCAATGCTTATGTTTGTAGAGCGTATATATAGATTT 349
DB 53 lleulleCysCys-----PheilleleuGluAsnillePheVal----- 66

QY 350 GGAGCGTCATCGCTTCGGGAGCTGTTCTTCTGAATCTTTTGTCTATCTGGCCAGCT--- 406
DB 67 -----LeuLeuThrlleTrpLysThrls 74

QY 407 -----CTACAGTATTGGATGCGCCATCCTTGTCTAGGACTCGCTGGG 451
DB 75 LysPheHisArgPrometTyrrPheilleGlyAsnLeuAlaLeuSerAspLeuLeuAla 94

QY 452 ATTGTTCTTGTATGCGGATGCGGTAGATCAAAATGTTCTGTATTTCGAAGAATCCGA 511
DB 95 -----GlyValAlaTyrrThrAlaAsnLeuLeu----- 104

QY 512 GAGGAATTTTATTGCTCTCAAGTCTTAAAAAATCTGAGAAAAAGGATATACCAAGGCT 571
DB 104 ----- 104

QY 572 TTGGAGCCATTTTGTATCTTAACATGTGACTACAGTATTGGCTCAGCACTTCTTTTCTTC 631
DB 105 -----SerGlyAlaThrThrTyrrLysLeuThrProAlaGlnTrpPhe 118

QY 632 CTAGATACAGGCGCTATTAAAGGTTTCTTGTGACATTTAGTAATTTCTCTTCA 691
DB 119 LeuArgGluGly-----SerMetPheValAlaLeuSerAlaSer 131

QY 692 ATGTTTACG-----GCTCTTTTCATGACTAAATTTTCTCATCTCTGATG----- 739
DB 132 ValPheSerLeuLeuAlaIleAlaIleGluArgTyrrIleThrMetLeuLysMetLysLeu 151

QY 740 ---AATAAGACCCAAATACACAGTTGATGATGAATAAGTTCTGCTGGGATAAAGCAT 796
DB 152 HisAsnGlySerAsnAsnPheArgLeuPheLeuLeu----- 163

QY 797 GATTTCTTGAGAGGATGCAAAAACCTTTGGGCTGTTTCTGGNAGTGTCTTTCTTTAGGT 856
DB 164 -----IleSerAlaCys-----TrpValIleSer-----LeuIleLeuGly 175

QY 857 TCGGTTGCTCTCGGTTGGAGCTGGAATCCGTTTGGGAATGCAATTTAAAGGAGG 916
DB 176 -----GlyLeuProIleMetGlyTrpAsnCysIle----- 185

QY 917 TATGCTTTTACCTTTAATCCAAAAGACATGGCATCAGCGATGTTGCTCAAAATGCTGGC 976
DB 185 ----- 185

QY 977 AAAGTTGTGCATAAACTACAGAACTGGTCTTTCTTCTAGAGACTTCGTTATTCAAACA 1036
DB 186 -----SerAlaLeuSerSer----- 190

QY 1037 TTGGATCTTCAGAAAAGATCAAAATCTATTTTGTAGTAAAGCTTTAAGCTATATAAG 1096
DB 191 -----CysSerThrValLeuProLeuTyrrHisLysHisTyrrIleLeuPheCysThrThr 208

QY 1097 CAGATACAGCGCTCTCTCTAAATAATACGATCATGAGCTGGCGTTATTGTGGGATT--- 1153
DB 209 ValPheThrLeuLeuLeuSerIleValIleLeu-----TyrrCysArgIleTyrr 225

QY 1154 ---GTTGTCAAGAACAGGCTTAGATTTCTTACGGAATACTTAAACGAAACGCAAAATTT 1210
DB 226 SerLeuValArgThrArgSerArgLeu-----ThrPheArgLysAsnilleSerLys 243

QY 1211 TGGTCAAAAGGTAAGCAGCAAACTATCGAAGAAATCGCTTATCAGCGCACCATCGGGCTT 1270
DB 244 AlaSerArgSerSerGluLys-----SerLeuAlaLeu 254

QY 1271 TTAGGAGCTTTGGCAATCATCTTG---CTCTATGTAGTTTGGCTTTGAATGGCAATAT 1327
DB 255 LeuLysThrValIleIleValLeuSerValPheIleAlaCys-----TrpAlaPro 271

QY 1328 GCTTTTACGTCCGATGCGCTTTAAATTTATCATGACCTTTTGGCTACCTGTGAGTCTTCTTT 1387
DB 272 LeuPheilleleuLeuLeuAspValGlyCysAllyValLysThrCysAspIleLeuPhe 291

QY 1388 ATAGCACATTTCTTTTGAAGAAAATTCAAATAGATTTGCAAGCCATTGGTGGCTTTAATG 1447
DB 292 ArgAlaGluTyrrPheLeu-----Valleu 299

QY 1448 ACTGTATTGGGTATTTCATTAACAATACATCTTTGATCATTTTGTATGCTATTCGTGAAGAT 1507
DB 300 AlaValLeuAsnSerGlyThrAsnProIleIle----- 310

QY 1508 CGCCAAGCGAACCTGTTTACCCTCATGCTGTTTAACTTAATGATGCCCTTCAAAGACG 1567
DB 311 -----TyrrThrLeuThrAsnLysGluMetArgArgAla 321

QY 1568 TTCAGCCGC 1576
DB 322 PheIleArg 324

RESULT 41
US-10-037-616-19
; Sequence 19, Application US/10037616
; Patent No. US20020123148A1
; GENERAL INFORMATION:
; APPLICANT: English, Denis
; APPLICANT: Kovacs, Richard J.
; APPLICANT: Rizzo, Maria T.
; APPLICANT: Sliva, Daniel T.
; TITLE OF INVENTION: Sphingolipid Compositions and Methods for Their Therapeutic Use
; FILE REFERENCE: 7042-119
; CURRENT APPLICATION NUMBER: US/10/037,616
; CURRENT FILING DATE: 2002-04-19
; PRIOR APPLICATION NUMBER: US 60/243,887
; PRIOR FILING DATE: 2000-10-27
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 19
; LENGTH: 381
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-037-616-19

Alignment Scores:
Pred. No.: 0.117 Length: 381
Score: 96.50 Matches: 83
Percent Similarity: 32.05% Conservative: 59
Best Local Similarity: 18.74% Mismatches: 116
Query Match: 2.91% Indels: 185
DB: 12 Gaps: 22

US-09-868-987-1 (1-1864) x US-10-037-616-19 (1-381)

QY 290 ATCTCAGCATCTGGCTTGGCAATGCTTATGTTTGTAGAGCGTATATATAGATTT 349
DB 53 lleulleCysCys-----PheilleleuGluAsnillePheVal----- 66

QY 350 GGAGCGTCATCGCTTCGGGAGCTGTTCTTCTGAATCTTTTGTCTATCTGGCCAGCT--- 406
DB 67 -----LeuLeuThrlleTrpLysThrls 74

QY 407 -----CTACAGTATTGGATGCGCCATCCTTGTCTAGGACTCGCTGGG 451
DB 75 LysPheHisArgPrometTyrrPheilleGlyAsnLeuAlaLeuSerAspLeuLeuAla 94

QY 452 ATTGTTCTTGTATGCGGATGCGGTAGATCAAAATGTTCTGTATTTCGAAGAATCCGA 511
DB 95 -----GlyValAlaTyrrThrAlaAsnLeuLeu----- 104

QY 512 GAGAAATTTTATGTCTCAAGCTTTAAATAATCTGTAGAAAAGATATACCAAGCT 571
 Db 104 ----- 104
 QY 572 TTTGAGCCATTTTGAATCTAAGTACATGATTTGGCTGACACTTTCTTCTTC 631
 Db 105 -----SerGlyAlaThrThrIyrIysLeuThrProIaGlnTrpHe 118
 QY 632 CTAGATACAGGCGCTTAATAAGGTTGGCTTGAACATGATTTTGAAGATTTCTCTCA 691
 Db 119 LeuArgIuGly-----SerMetPheValAlaLeuSerAlaSer 131
 QY 692 ATGTTTACG-----GCTCTTTTCATGACTAAATTTTCTTCATGCTGGATG----- 739
 Db 132 ValPheSerLeuLeuAlaIleAlaIleGluArgIyrIleThrMetLeuLysMetLysLeu 151
 QY 740 ---AATAAGCCCAACATACACAGTGCATGATGATGATTAAGTTCGGGGATTAACAT 796
 Db 152 HisasnGlySerAsnAsnPheArgLeuPheLeu----- 163
 QY 797 GATTTCTTGAGAGATGCAAAAACTTTGGCTGTTTCTGAGAGTGTTTTCTTTAGT 856
 Db 164 -----IleSerAlaCys-----TrpValIleSer-----LeuIleLeuGly 175
 QY 857 TGCCTTGTCTCGGGTTTGGAGCTGGAATTCGGTTTGGGAATGATTTTAAAGAGG 916
 Db 176 -----GlyLeuProIleMetGlyTrpAsnCysIle----- 185
 QY 917 TATGCTTACCTTTAATCCAAAAGAGATGCATACGATGTGTCAATGCGTGGC 976
 Db 185 ----- 185
 QY 977 AAAGTTGTGATAACTACAGAAAGCTGCTTTCTTCTAGACCTTCGATTCAACA 1036
 Db 186 -----SerAlaLeuSer----- 190
 QY 1037 TTTGATCTTCAAGAAAGATCAAAATCTATTAGTGAATAAGCTTAAAGTATATAG 1096
 Db 191 -----CysSerThrValLeuProLeuIyrHisLysHisIyrIleLeuPheCysThrThr 208
 QY 1097 CAGATACAGGCTCTCTCTAAATTAACATGATGATGAGCTGGCTTATTTGGAGATT-- 1153
 Db 209 ValPheThrLeuLeuLeuLeuSerIleValIleLeu-----TyrCysArgIleIyr 225
 QY 1154 ---GTTGCAAGAAACAGCCCTAGATTCTCTACGAAAACCTTAAACGAAACGAAATTT 1210
 Db 226 SerLeuValAlaGthrArgSerArgArgLeu-----ThrPheArgLysAsnIleSerLys 243
 QY 1211 TGTCAAAAGGTACGACGAACCTATCGAAGAAATGCTTACAGGCGACCATGGGCTT 1270
 Db 244 AlaSerArgSerSerGluLys-----SerLeuAlaLeu 254
 QY 1271 TTAGAGCTTTGGCAATCATCTTG--CTATATGAGTTTGGCTTGAATGCAATAT 1327
 Db 255 LeuIyrThrValIleIleValLeuSerValPheIleAlaCys-----TyrAlaPro 271
 QY 1328 GCTTTCAGTGGCGTATGCGCTTATTAATCATGACCTTTGGTACCTGTCAGCTTGT 1387
 Db 272 LeuHeIleLeuLeuLeuLeuAspValGlyCysLysValLysThrCysAspIleLeuPhe 291
 QY 1388 ATGACACTTTCTTTTGAAGAAAATTCAAATAGTTTGCAGAGCATTTGGCTTTATG 1447
 Db 292 ArgIaGluIyrPheLeu-----ValLeu 299
 QY 1448 ACTGATTTGGGGTATTTCAATAACATATCTTGATCATTTTGTATTCGTAAGAT 1507
 Db 300 AlaValLeuAsnSerGlyThrAsnProIleIle----- 310
 QY 1508 CGCCAAAGCAACCTGTATACCCCTATGACATGTTTATGATGAGCCCTCAAAAGAGC 1567
 Db 311 -----TyrThrLeuThrAsnLysGluMetArgArgAla 321
 QY 1568 TTCAGCCGC 1576

Db 322 PheIleArg 324
 RESULT 42
 US-09-971-228-5
 / Sequence 5, Application US/09971228
 / Patent No. US2002015512A1
 / GENERAL INFORMATION:
 / APPLICANT: Liao, X. Charlene
 / APPLICANT: Masuda, Esteban
 / APPLICANT: Chu, Peter
 / APPLICANT: Pardo, Jorge
 / APPLICANT: Li, Congfen
 / APPLICANT: Zhao, Haoran
 / APPLICANT: Jiang, Yingping
 / APPLICANT: Rigel, Pharmaceuticals, Incorporated
 / TITLE OF INVENTION: EDG: Modulators of Lymphocyte Activation and Migration
 / FILE REFERENCE: 021044-000310US
 / CURRENT APPLICATION NUMBER: US/09/971,228
 / PRIOR FILING DATE: 2002-04-19
 / PRIOR APPLICATION NUMBER: US 60/284,763
 / NUMBER OF SEQ ID NOS: 15
 / SOFTWARE: PatentIn Ver. 2.1
 / SEQ ID NO 5
 / LENGTH: 382
 / TYPE: PRT
 / ORGANISM: Homo sapiens
 / FEATURE:
 / OTHER INFORMATION: wild-type human endothelial differentiation
 / OTHER INFORMATION: G-protein coupled receptor (GPCR) 1 (EDG1)
 US-09-971-228-5
 Alignment Scores:
 Pred. No.: 0.117 Length: 382
 Score: 96.50 Matches: 83
 Percent Similarity: 32.05% Conservative: 59
 Best Local Similarity: 18.74% Mismatches: 116
 Query Match: 2.91% Indels: 185
 DB: 9 Gaps: 22
 US-09-868-987-1 (1-1864) x US-09-971-228-5 (1-382)
 QY 290 ATCTCAGACATGCTGCTGGCTTGAATGCTTATTTGATGAGCTATATTATGATTT 349
 Db 53 IleLeuIleCysCys-----PheIleIleLeuGluAsnIlePheVal----- 66
 QY 350 GAGAGGCTATCGCTTCGGAGCGTTCCTGATCTTTGCTTATTCGAGGAGCT--- 406
 Db 67 -----LeuLeuThrIleTrpLys 74
 QY 407 -----CTACAGATTTTGGATGCGCACTGCTTGCAGACTCGCTGGG 451
 Db 75 LysPheHisArgProMetTyrTrpPheIleGlyAsnLeuAlaLeuSerAspLeuAla 94
 QY 452 ATTGTTCTTCTATGCGGATGCGCTAGATGCAAAATTTTGTATTCGAAAGATCCGA 511
 Db 95 -----GlyValAlaIyrThrAlaAsnLeuLeu----- 104
 QY 512 GAGAAATTTTATGTCTCAAGCTTTAAATAATCTGTAGAAAAGATATACCAAGCT 571
 Db 104 ----- 104
 QY 572 TTTGAGCCATTTTGAATCTAAGTACATGATTTGGCTGACACTTTCTTCTTC 631
 Db 105 -----SerGlyAlaThrThrIyrIysLeuThrProIaGlnTrpPhe 118
 QY 632 CTAGATACAGGCGCTTAATAAGGTTGGCTTGAACATGATTTTGAAGATTTCTCTCA 691
 Db 119 LeuArgIuGly-----SerMetPheValAlaLeuSerAlaSer 131
 QY 692 ATGTTTACG-----GCTCTTTTCATGACTAAATTTTCTTCATGCTGGATG----- 739


```
QY 1133 ACCTGGCGTTATTGTGGGATTTGTTGTGAGAAACAGCGCTAGATTCTCTACGGAAACTCT 1192
Db 296 PheLeuAlaTyrPheGly----- 301
QY 1193 AAACGAAACGCAAAATTTTGGTCAAAGGTAAGCAGCAAACTATCGAAGAAATG----- 1246
Db 302 -----ValSerSerAlaLeuThrLeuMetMetProTyr 312
QY 1246 ----- 1246
Db 313 TyrLysLeuGlnProGluSerProLeuProGluAlaPheThrTyrValGlyTrpGluPro 332
QY 1247 ----CCTTATCAGCGACACCTCGGCTTTTAGGAGCTTTGGCAATCATCTTGCTC----- 1297
Db 333 AlaArgTyrLeuValAlaIleGlySerLeuCysAlaLeuSerThrSerLeuLeuGlySer 352
QY 1298 ---TATGTGAGTTTGGCTTTGAATGGCAATATGCT-----TTCAAGTCCGCTA 1342
Db 353 MetPheProMetProArgValIleTyrAlaMetAlaGluAspGlyLeuLeuPheArgVal 372
QY 1343 TGCCTTTAATTATCATGAC-----CTTTTGGCTACCTGTGCA----- 1378
Db 373 LeuAlaArgValHisAsnGlyThrHisThrProIleValAlaThrValValSerGlyVal 392
QY 1379 ---GTCTTGTATTAGCACATTTCTTTTGAAGAAATTCAAATAGATTGCAAGCCATT 1435
Db 393 IleAlaAlaPheMetAlaPheLeuPheGluLeuThrAspLeuValAspLeuMetSerIle 412
QY 1436 GTGCTTTAATGACTGTATTGGGTATTTCATTAAACAATCTTTGATCATTTTGTATCGT 1495
Db 413 Gly-----ThrLeuAlaTyrSerLeuValSerIleCysValLeuIleLeuArg 429
QY 1496 ATTCGTGAAGATCGC----- 1510
Db 430 TyrGlnProAspGlnGluMetLysAsnGlyGluGluValGluLeuGlnGluArg 449
QY 1511 ----CAAGCGAACTGTTTACCCTATGCAATGTTTGTAGTTAAT---GATCCCTTCAA 1561
Db 450 ThrLeuGluAlaGluLysLeuThrValGlnAlaLeuPheCysGlnValAspSerIlePro 469
QY 1562 AAGAGCTTCAGC---CGCAGCGTAATGACACAGCTACAACTCTATCAGTTTGTGTAATG 1618
Db 470 ThrLeuLeuSerGlyArgIleValTyrValCysSerSerLeuAlaValLeuLeuThr 489
QY 1619 CTTTGTGTTTATA-----GGCGGCTCCTCTGCTTT 1648
Db 490 ValLeuCysLeuValLeuThrTrpTrpThrProLeuHisSerGlyAspProValTrp 509
QY 1649 AATTTGCACTTTATTATGACCATAGGATTTCTTAGGAACCTTTATCGTCTCTTTATATT 1708
Db 510 ValThrValValLeuLeuLeuGlyLeuIleLeuGlyIleSerGlyValIleTrpArg 529
QY 1709 GCACCA 1714
Db 530 GlnPro 531
RESULT 45
US-09-971-228-14
; Sequence 14, Application US/09971228
; Patent No. US2002015512A1
; GENERAL INFORMATION:
; APPLICANT: Liao, X. Charlene
; APPLICANT: Masuda, Eateban
; APPLICANT: Chu, Peter
; APPLICANT: Pardo, Jorge
; APPLICANT: Li, Congfen
; APPLICANT: Zhao, Haoan
; APPLICANT: Jiang, Yingping
; APPLICANT: Rigel Pharmaceuticals, Incorporated
; TITLE OF INVENTION: EDG: Modulators of Lymphocyte Activation and Migration
; FILE REFERENCE: 021044-000310US
; CURRENT APPLICATION NUMBER: US/09/971,228
```

```
; CURRENT FILING DATE: 2002-04-19
; PRIOR APPLICATION NUMBER: US 60/284,763
; PRIOR FILING DATE: 2001-04-18
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 14
; LENGTH: 314
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: mutant #2, C-terminally truncated variant of
; OTHER INFORMATION: human endothelial differentiation G-protein
; OTHER INFORMATION: coupled receptor (GPCR) 1 (EDG1)
US-09-971-228-14
Alignment Scores:
Pred. No.: 0.173 Length: 314
Score: 94.50 Matches: 76
Percent Similarity: 33.16% Conservative: 52
Best Local Similarity: 19.69% Mismatches: 103
Query Match: 2.85% Indels: 155
DB: 9 Gaps: 20
US-09-868-987-1 (1-1864) x US-09-971-228-14 (1-314)
QY 290 ATCTCAGCATCTGTGGCTTGGCAATGCTTATGTTTGTATGATGAGCGTATATTATAGATT 349
Db 53 IleLeuIleCysCys-----PheIleIleLeuGluAsnIlePheVal----- 66
QY 350 GGAGCGTCATCGTCTCGGAGCTGTTCTTCTGAATCTTTTGTCTTATCTGGCAGCT--- 406
Db 67 -----LeuLeuThrIleTrpLysThrLys 74
QY 407 -----CTACAGTATTGGATGGCCACTCACCTTGTCTCAGGACTCGCTGG 451
Db 75 LysPheHisArgProMetTyrTyrPheIleGlyAsnLeuAlaLeuSerAspLeuLeuAla 94
QY 452 ATTGTCTTGTCTAGGAGTGGCGGTAGATGCAAAATGTTCTTGTATTTCGAAAGAAATCCGA 511
Db 95 -----GlyValAlaTyrThrAlaAsnLeuLeuLeu----- 104
QY 512 GAGGAATTTTATTGTCTCAAAGTCTTAAANAATCTGTAGAAAAGGATATACCAGGCT 571
Db 104 ----- 104
QY 572 TTGAGGCCATTTTGTATTCTAATCTGACTACAGTATTGGCTCAGCAGCTTCTTTTCTTC 631
Db 105 -----SerGlyAlaThrThrTyrLysLeuThrProAlaGlnTrpPhe 118
QY 632 CTAGATACAGGCGCTATTAAAGGGTTTGTCTTGACATTGATTTTAGGAATTTTCTCTTCA 691
Db 119 LeuArgGluGly-----SerMetPheValAlaLeuSerAlaSer 131
QY 692 ATGTTTAGC-----GCTCTTTTCTATGACTAAATTTTCTTCATGCTGTGGATG----- 739
Db 132 ValPheSerLeuLeuAlaIleAlaIleGluArgTyrIleThrMetLeuLysMetLysLeu 151
QY 740 ---AATAAGACCCACATACACAGTTGCATATGATGAATAAGTTCTGCGGGATAAAGCAT 796
Db 152 HisAsnGlySerAsnAsnPheArgLeuPheLeuLeu----- 163
QY 797 GATTTCTTGAGAGGATGCAAAAAAATTTTGGGCTGTCTTCTGGAAGTGTCTTTCTTTTAGGT 856
Db 164 -----IleSerAlaCys-----TrpValIleSer-----LeuIleLeuGly 175
QY 857 TCGGTGCTCTCGGGTTTGGAGCCTGGAATTCGGTTTGGGAATGGATTTTAAAGGAGGG 916
Db 176 -----GlyLeuProIleMetGlyTrpAsnCysIle----- 185
QY 917 TATGCTTTACTTTTATCCAAAAGAGCATGGCATCGCGATGTGTCTCAATGCGTGGC 976
Db 185 ----- 185
```

```
QY 977 AAAGTTGATGAATAACTACAGAGAGCTGGCTTTCTCTAGAGACTCCGATATCAACA 1036
Db 186 -----SerAlaLeuSer----- 190
QY 1037 TTGGATCTTCAGAAAAGATCAAAATCTATTATTAGTATAAGCTTAAAGCTATACTAAG 1096
Db 191 -----CysSerThrValLeuProLeuTyrHisIysHisTyrIleLeuPheCysThrThr 208
QY 1097 CAGATACAGAGCTCTCTCTAAATTAACGATCATGAGCTGGCGTTAATGTGGATT--- 1153
Db 209 ValPheThrLeuLeuLeuSerIleValIleLeu-----TyrCysArgIleTyr 225
QY 1154 ---GTTGTCAGAAACAGGCCCTAGATTCTCTACGGAACCTAAACGAACGCAAAATTT 1210
Db 226 SerLeuValArgThrArgSerArgArgLeu-----ThrPheArgLysAsnIleSerLys 243
QY 1211 TGGTCAAGGTAAAGCAGCAAACTATCGAAGAAATGCGTTATCAGCGACCATCGGCTT 1270
Db 244 AlaSerArgSerSerGluLys-----SerLeuAlaLeu 254
QY 1271 TTAGAGCTTTGGCAATCATCTTG--CTCATGTGAGTTTGGCGTTGAATGGCAATAT 1327
Db 255 LeuLysThrValIleIleValLeuSerValPheIleAlaCys-----TrpAlaPro 271
QY 1328 GCTTTCAGTCCGATATGCGCTTTAATCATGACCTTTGGCTACCTGTGCAGTCTTGT 1387
Db 272 LeuPheIleLeuLeuLeuLeuAspValGlyCysLysValIlySThrCysAspIleLeuPhe 291
QY 1388 ATAGCACATTTCTTTTG 1405
Db 292 ArgAlaGluTyrPheLeu 297
```

Search completed: December 17, 2002, 12:50:03
Job time : 89.5 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 17, 2002, 06:25:53 ; Search time 1976 Seconds
(without alignments)
15277.524 Million cell updates/sec

Title: US-09-868-987-1
Perfect score: 1864
Sequence: 1 atggacttcgcattgtc.....atgacaattcagataatgc 1864

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

```
EST.*
1: em_estba.*
2: em_esthum.*
3: em_estin.*
4: em_estmu.*
5: em_estov.*
6: em_estpl.*
7: em_estro.*
8: em_hcc.*
9: gb_est1.*
10: gb_est2.*
11: gb_hcc.*
12: gb_est3.*
13: gb_est4.*
14: gb_est5.*
15: em_estfun.*
16: em_estom.*
17: gb_gss.*
18: em_gss_hum.*
19: em_gss_inv.*
20: em_gss_pln.*
21: em_gss_vrt.*
22: em_gss_fun.*
23: em_gss_mam.*
24: em_gss_mus.*
25: em_gss_other.*
26: em_gss_pro.*
27: em_gss_rod.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	55.8	3.0	1101	17	CNS00LT2
2	54.4	2.9	187	17	BH371348
3	51.8	2.8	602	17	BH182894
4	51.8	2.8	602	17	CNS0708R
5	50.8	2.7	731	17	BH397261
6	49.2	2.6	1101	17	CNS00FOO

7	48	2.6	1101	17	CNS0039G
8	47.2	2.5	1101	17	CNS0100X
9	46.8	2.5	1101	17	CNS000D1
C 10	46.6	2.5	685	13	BJ353037
C 11	45.4	2.4	1101	17	CNS00807
C 12	45.2	2.4	963	9	AL566565
C 13	45.2	2.4	988	17	CNS06PBX
14	44.4	2.4	516	10	BE118484
15	43.8	2.3	548	17	AZ303132
C 16	43.8	2.3	1125	9	AL547503
17	43.6	2.3	417	14	BQ399330
18	43.6	2.3	500	9	AU087215
C 19	43.6	2.3	953	17	CNS02701
20	43.4	2.3	1101	17	CNS00HFX
C 21	43.2	2.3	803	17	CNS00DM2
C 22	42.8	2.3	863	17	AZ550729
C 23	42.8	2.3	332	12	BG350812
C 24	42.6	2.3	1001	17	CNS0155H
C 25	42.4	2.3	374	14	T72643
C 26	42.4	2.3	713	17	B80295
C 27	42.4	2.3	878	17	CNS0187R
C 28	42.4	2.3	885	17	CNS0187R
C 29	42.2	2.3	1101	17	CNS0039G
C 30	41.8	2.2	946	17	CNS005V6
31	41.8	2.2	976	17	CNS04ESM
C 32	41.8	2.2	1096	17	CNS051P2
C 33	41.8	2.2	1101	17	CNS003B4
C 34	41.6	2.2	930	17	AZ680882
35	41.6	2.2	931	17	AZ541040
36	41.6	2.2	1080	17	CNS008PP
37	41.4	2.2	552	12	BF279297
C 38	41.4	2.2	676	13	BJ349244
C 39	41.4	2.2	924	12	BG576544
40	41.4	2.2	1101	17	CNS008N7
C 41	41.4	2.2	1101	17	CNS001Q0
C 42	41.2	2.2	546	17	AZ036289
C 43	41.2	2.2	629	12	BF295860
C 44	41.2	2.2	750	17	CNS0111D
C 45	41	2.2	436	14	C92475

ALIGNMENTS

CNS00LT2 1101 bp DNA linear GSS 14-JUN-1999
Drosophila melanogaster genome survey sequence TET3 end of BAC:
BACR48P19 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.

AL078714

AL078714.1 GI:5102004

GSS.

Drosophila melanogaster.

Drosophila melanogaster.

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 1101)

Genoscope.

Direct Submission

Submitted (11-JUN-1999) Genoscope - Centre National de Sequençage :

BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr

- Web : www.genoscope.cns.fr)

Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP).

The BDGP is constructing a physical map of the Drosophila

melanogaster genome using these BACs. For further information

please see <http://www.fruitfly.org> The BDGP Drosophila

melanogaster BAC library was prepared by Kazutoyo Onoe and

Aaron Mammose in Pieter de Jong's laboratory in the Department of

Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,

NY. The library is named RPCI-98 and was constructed by partial

ECORI digestion of *Drosophila* DNA provided by the BDGP from the isogenic strain y2; cn bw sp; the same strain used for the BDP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES
Source
1. 1101
Location/Qualifiers
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="BACR48P19"
/clone_lib="RPCI-98"
/note="end : TERT"

BASE COUNT 469 a 6 c 69 g 151 t 406 others
ORIGIN

Query Match 3.0%; Score 55.8; DB 17; Length 1101;
Best Local Similarity 21.9%; Pred. No. 0.002;
Matches 117; Conservative 165; Mismatches 253; Indels 0; Gaps 0;

Qy 539 AAAAAATCTAGAAAAGATATACCAAGCCTTTGGAGCCATTGATTTAACTTG 598
Db 463 AAAAAAAAAAAAAAAAAATTTTWTWKATTTTDDADDKAAAWMD 522
Qy 599 ACTACAGATTGGCCTGACCTCTTTCTCCAGATACAGGCCCATTAAGGTTT 658
Db 523 TTTTAAATTTTAAATTAATTTTCTTWTWKATKDKWTKKTAATAAAMWTTT 582
Qy 659 GCTTGCATGATTTTGAATTTCTCTCAATGTTTACGGCTTTTCATGACTAA 718
Db 583 TTTAAATTTTCTTAAATTAATTTTCTTAAATTTTAAATTTTAAATTTTAA 642
Qy 719 TTTTCTTCAATGCTGTGATGAATTAAGCCAACTACACAGTTGCATATGATTAAG 778
Db 643 TTTAAATTTTCTTAAATTTTCTTAAATTTTAAATTTTAAATTTTAAATTTTAA 702
Qy 779 TTCGTGGGATTAACAGATTTCTTGAGAGATGCAAAACTTTGGCGTTTCGGA 838
Db 703 KKKKAAADKKDKAKWGAATKTKKDKKAAWAAADKDKKGGKKKKGGKKGG 762
Qy 839 AGTGTCTTCTTAAAGTGTGCTGCTCGGCTTTGGAGCCTGAATTCGTTTGGGA 898
Db 763 KKKKKKKGGGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGG 822
Qy 899 ATGATTTTAAAGAGGCTATGCTTTAATCCAAAGAGCATGACATGCGAT 958
Db 823 ADRTKKTWDAAAAAAAKKTDKGKKKKTKTKKKKKKKKKKKGGKDDAAAKKKK 882
Qy 959 GTTGCTCAATGCGTGGCAAGGTGTGATTAACATACAGGAAGCTGCTTTCTTGA 1018
Db 883 GTKKKKGGGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGG 942
Qy 1019 GACTTCGATTTCAAACTTTGATCTTCAGAAAGATCAAAATCTATTTATG 1073
Db 943 GKKTTKTKKKKKKKKDAAGKTKKRAAADAAADMTDAATTKKKKKKKDKK 997

RESULT 2
BH371348 187 bp DNA linear GSS 10-DEC-2001
LOCUS BH371348
DEFINITION AG-ND-132P6.TR ND-TM Anopheles gambiae genomic clone AG-ND-132P6,
DNA sequence.
ACCESSION BH371348
VERSION BH371348.1 GI:17317473
KEYWORDS GSS.
SOURCE African malaria mosquito.
ORGANISM Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;
Anopheles.
REFERENCE 1 (bases 1 to 187)
AUTHORS Shetty,J., Malek,J., Koo,H., Collins,F., Gardner,M. and Loftus,B.J.
TITLE Direct Submission of BAC-end sequences from *Anopheles gambiae*

JOURNAL

Unpublished (2001)
Other_GSSs: AG-ND-132P6.TF

COMMENT
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: bjl@fugate.org

This clone is from an *A. gambiae* BAC library (ND-TM) provided by F.H. Collins and sequenced by The Institute for Genomic Research (TIGR). The BAC library was generated from *A. gambiae* PEST strain DNA. All DNA was extracted from newly hatched first instar larvae to minimize the inclusion of DNA from microorganisms that inhabit the gut. The DNA is derived from mixed sexes of larvae. The BAC library was constructed at Texas A&M University BAC Center University, College Station, Texas 77843-2123, USA using a HindIII partial digest.
Seg primer: M13 Rev
Class: BAC ends.

FEATURES
Source
1. 187
Location/Qualifiers
/organism="Anopheles gambiae"
/strain="PEST"
/db_xref="taxon:7165"
/clone="AG-ND-132P6"
/clone_lib="ND-TM"
/note="Vector: pECBAC1; Site_1: HindIII"

BASE COUNT 53 a 42 c 30 g 62 t
ORIGIN

Query Match 2.9%; Score 54.4; DB 17; Length 187;
Best Local Similarity 56.0%; Pred. No. 0.0025;
Matches 103; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

Qy 1440 CTTTATGACTGTATGGGATTTCAATTAACAATCTTGATCATTTTATGATGATTC 1499
Db 2 CTGATTAACAGATTATGATATTTCTATTAACATACGATATGATTTGACCGATCC 61
Qy 1500 GTGAAGATGCCCAAGCAACCTGTTTACCCCTATGATGTTTATGATATGATGCCCTTC 1559
Db 62 GTGAATACCTTACGAAAGAAATCTATCACTTCGAGGCTGTTGATGATTCATCA 121
Qy 1560 AAAAGAGCTTACGCCGACGATATGACACAGTCAACTCTATCATGTTTGTATATGC 1619
Db 122 CCAGCAGGTTGGTAGAATCACTTCATTCATCACTACCATATGTTATCCTTGGCA 181
Qy 1620 TTTT 1623
Db 182 TCTT 185

RESULT 3
BH182894 602 bp DNA linear GSS 19-OCT-2001
LOCUS BH182894
DEFINITION 022_B_07-rev SmbAC1 Schistosoma mansoni genomic clone 022B07 5',
DNA sequence.
ACCESSION BH182894
VERSION BH182894.1 GI:16287528
KEYWORDS GSS.
SOURCE Schistosoma mansoni.
ORGANISM Schistosoma mansoni.

Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;
Strigeidae; Schistosomatidae; Schistosomatidae; Schistosoma.
REFERENCE 1 (bases 1 to 602)
AUTHORS Le Paslier,M.-C., Pierce,R.J., Merlin,F., Hirai,H., Wu,W., Williams
D.L., Johnston,D., Loverde,P.T. and Le Paslier,D.
TITLE Construction and characterization of a Schistosoma mansoni
bacterial artificial chromosome library
JOURNAL Genomics 65 (2), 87-94 (2000)
MEDLINE 20247247
COMMENT Other GSSs: 022_B_07-21
Contact: Pierce_RJ

INSERM U 167
Institut Pasteur de Lille
1 rue du Professeur A. Calmette, 59019-Lille, France
Tel: (33) (0)3 20877783
Fax: (33) (0)3 20877888
Email: Raymond.Pierce@pasteur-lille.fr
CNS sequencing ID=DG0AA022CA04BFI
Plate: 022 row: B column: 07
Seq primer: M13 reverse primer
Class: BAC ends
High quality sequence stop: 602.
Location/Qualifiers
1..602
/organism="Schistosoma mansoni"
/strain="Puerto-Rican"
/db_xref="taxon:6183"
/clone="022B07"
/clone_lib="SmbAC1"
/sex="mixed"
/dev_stages="cercariae"
/lab_host="Biomphalaria glabrata"
/note="Vector: pBelOBAC 11; Site 1: Hind III; Partially
Hind III digested and size-selected S. mansoni cercarial
DNA was ligated into Hind III digested pBelOBAC 11 vector
and used to transform E. coli DH10B. The complete library
contains 23808 clones from 4 independent
sizing-ligation-transformations. Average insert size
ranges from 70-127 kb and genome coverage is 7.9-fold."
BASE COUNT 194 a 88 c 121 g 158 t 41 others
ORIGIN
Query Match 2.8%; Score 51.8; DB 17; Length 602;
Best Local Similarity 42.7%; Pred. No. 0.018;
Matches 213; Conservative 5; Mismatches 274; Indels 7; Gaps 1
QY 255 CTCATCTTGGGAAAAACAATGACACAAGGCATTATCTCAGCATGCTGTGGCTTGGCAA 314
Db 10 CTTCCTTAGGACAAAAGCAATGTATGCAGGTATTTATCTTCATTTAGTATGGATCTTCT 69
QY 315 TGCATTATGTTTGTAGAGCGGTATATTATAGATTGAGAGCGGTATCGCTTCGGGAGCTG 374
Db 70 TWTATGCTGTGGATGGTGTCTATTACGGTAGAGCGGTTGGTATGCAAAATGCGGCTT 129
QY 375 TTCTTCTGATCTTTTGCTTATCTGGCGAGCTCTACAGTATTTGGATCGGCCCTACCT 434
Db 130 TATTATTAACAGTATTTATTTCTTATTGGTGTAAATGGCAAGTTTGGTTGTATTAAAT 189
QY 435 TGTCAAGACTTCGCTGGGATGTCTTTGTGTATGGGATGGCGTAGATGCAAAATGCTTCTTG 494
Db 190 TGCAGGTATTCAGGTATCGSATWAACATTAGGTACTGCAATTTGATGCGCAATCATTA 249
QY 495 TATTGGAAGATCCGAGAGGAATTTTATGCTCTCAAGTCTTTAAAAAATCTGTAGAAA 554
Db 250 AAAAGAAAGAGCAAAAGAGAAATWACGTGAAGGTAAAACTTTAGCAGAGNAGTTGACA 309
QY 555 AAGGATAT-----ACCAAGGCTTTTGAGCGCATTTTGTGATTTCTTAAGTACAGTA 607
Db 310 TTTCAATAGTTGGAAGAGGCGCANGGTAAACATAGTAGAAGCAACGTTACGCATATC 369
QY 608 TTGGCCTCAGCACTCTCTTTTCTCTCTAGATACAGGGCCTATTAAGAGGTTTGTCTTGACA 667
Db 370 TTAACGGNATCANCITTAANNNGCNTNGAACCGGACCCAGCAGGANTTTCGNNACA 429
QY 668 TTGATTTTAGGAATTTTCTCTTCAATGTTTACGGCTCTTTTCATGACTAAATTTTCTTC 727
Db 430 ANANCAAMAGGNNAAAAAANCNANANANANANANANANANANANANANANANANANANAN 489
QY 728 ATGCTGTGGATGAATAAGA 746
Db 490 GACAAAGAGATGCGGAAGA 508

CNS0706R	CNS0706R	T3 end of clone 022CA04 of Schistosoma mansoni, genomic survey sequence.	602 bp	DNA	linear	GSS 10-OCT-2001 from strain Puerto-Rican
LOCUS DEFINITION	Al619845	Al619845.1	GI:16033987			
ACCESSION	Al619845	Al619845.1	GI:16033987			
VERSION						
KEYWORDS						
SOURCE						
ORGANISM						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
MEDLINE						
PUBMED						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
COMMENT						
FEATURES						
source						
BASE COUNT						
ORIGIN						
Query Match						
Best Local Similarity						
Matches 213; Conservative						
QY	255	CTGATCTGGGAAAAACAATGTACACAGGCATTTCTCAGCAGTGTCTGGCTTGGCAA	314			
Db	10	CTTCTTTAGCACAAAAAGCAATTGATCAGGTATTTTATCTTCATTAGTTGGATTCTTCT	69			
QY	315	TGCTTATGTTTTGATCAGCGTATATTATAGATTGGAGCGTCATCGTTCGGGAGCTG	374			
Db	70	TWVATGTGTGTGGATGGTGTCTTATTACGGTAGAGCGGGTTGGTATGCAAATGCGGCTT	129			
QY	375	TTCTTCTGAATCTTTTGCTTATCTGGCGAGCTCTACAGTATTTGGATGCGCCACTCACCT	434			
Db	130	TATTATTAAAGTATATTCTTATTGGTGTAAATGGCAAGTTTGGTTTGTATTAAACAT	189			
QY	435	TGTCAGACTCGCTGGGATTGTTCTTGCTATGGGGATGGCGGTAGATGCAAATGTTCTTG	494			
Db	190	TGCCAGGTATTGCAGGTATCGSATWAAACATTTAGGTACTGCAAGTTTGATGCCAACATCATTA	249			
QY	495	TATTGGAAGAAATCCGAGAGGAATTTTTTATGTTCTCAAGTCTTTAAAAAATCTGTAGAA	554			
Db	250	AAAHGAAGAGCAAAAGAAGAATWACGTGAAGTGAATAAATCTTTAGCAGAGNAGTTGACA	309			
QY	555	AAGGATAT-----ACCAAGGCTTTTGGAGCCATTTTTTGATTCTTAACCTGACTACAGTA	607			
Db	310	TTTCAATGTTGGAAAGGAGCCANGCGTAACAATAGTAGAAGCAACGTTACCATATC	369			
QY	608	TGCGCTCAGACATTTCTTTTCTTCTTAGATACAGGGCCTATTAAAGGGTTGCTTTTGACA	667			
Db	370	TTAACGGGNATCANCTTTAAANNNGNTNGGAACCGGACCCANCCCAAGGANTTTCGNGNAACA	429			

RESULT 4


```
Db 896 KKAAKAKKAGGAGKADAKATAAAWAAKAKAKAADAADAAAADAAKADGTRWDGA 955
Qy 791 AAGCATGATTTCTTGAGAGGATGCAAAAACCTTTGGGCTGTTCTTGGAGGTGTTTCTT 850
Db 956 KAKRTKDAKARAKADAKADRAKKDKKKKKKKKKKKKKKKKKKKKKKKKKKKKKGA 1015
Qy 851 TTAGGTCGGTTCCTCGGTTTGGAGCCTGGGAATCCGTTTGGGAATGATTTAAA 910
Db 1016 KAKAKKATKKKKDKAKDKAKDKAKDKKKKKKKKKKKKKKKKKKKKKKKKKKK 1075
Qy 911 GGAGGGTAT 919
Db 1076 KKWKATATK 1084

RESULT 7
CNS0039G CNS0039G 1101 bp DNA linear GSS 03-JUN-1999
LOCUS Drosophila melanogaster genome survey sequence TET3 end of BAC #
DEFINITION BACR08K10 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL063921
VERSION AL063921.1 GI:4941778
KEYWORDS GSS.
SOURCE Drosophila melanogaster.
ORGANISM Drosophila melanogaster.
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS melanogaster BAC library was prepared by Kazutoyo Osoegawa and
TITLE Aaron Mammoser in Pieter de Jong's laboratory in the Department of
JOURNAL Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
COMMENT The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see http://www.fruitfly.org The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osoegawa and
Aaron Mammoser in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain y2; cn bw sp; the same strain used for the BDGP's
p1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES
source 1..1101
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="BACR08K10"
/clone_lib="RPCI-98"
/notes="end : TET3"
BASE COUNT 201 a 64 c 131 g 202 t 503 others
ORIGIN
Query Match 2.6%; Score 48; DB 17; Length 1101;
Best Local Similarity 21.0%; Pred. No. 0.22;
Matches 98; Conservative 161; Mismatches 207; Indels 0; Gaps 0;
Qy 1397 TTCCTTTTGAAGAAATTCAAATAGATTGCAAGCATTGGTCTTTTAATGACTGATTG 1456
Db 403 WWWWWTTTTTAAWAAAAAATAATTTTAAWAAAAAATAATTTTAAWAAAAAATAAT 462
Qy 1457 GGGTATCATTAACAATCTTGTATCATTTTGTATCGTATTTCGTGAAGATCGCAAGCG 1516
Db 463 WTTTAAWAAAAAATAATTTTAAWAAAAAATAATTTTAAWAAAAAATAATTTTAAWAAAA 522
```

```
Qy 1517 AACCTGTTTACCCCTATGCATGTTTATGTAATGATGCCCTTCAAAAGACGTTTCAGCGCG 1576
Db 523 AAAAAAATAAATAAATAATTTTAAWAAAAAATAATTTTAAWAAAAAATAATTTTAAWAAAA 582
Qy 1577 ACGTAATGACAAAGCTACAACTCTATCAGTTTGTGTAATGCTTTTGTATAGGCGCG 1636
Db 583 YTYWYTWTHYTHYTHYTHYTHYTHYTHYTHYTHYTHYTHYTHYTHYTHYTHYTHYTHY 642
Qy 1637 TCCTCTGCTTTTAAATTTTGCATTTTATGACCATAGGGGATTTCTTCTAGGAACCTTTATCG 1696
Db 643 YTCMYVYHYHMHHAHAHAHAHAHAHAHAHAHAHAHAHAHAHAHAHAHAHAHAHAHA 702
Qy 1697 TCTCTTATATGACCACTCTCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 1756
Db 703 YXXHHYTAHHTHTHHYAHYHYMYMYMYMYMYMYMYMYMYMYMYMYMYMYMYMYMYMY 762
Qy 1757 TAAGTACCGTTAAACTTAATCTAAGCTGTAGCAATATAAAATCTCTTCTGGGACTTTAG 1816
Db 763 AMWVHHHAHYAAAAAATAATTHHHYHHTTHYHMYHYMYHYMYHYMYHYMYHYMYHYMY 822
Qy 1817 TCCCAAGGCGCCCTGCTGCTATTAATTAATGACAAATTCAGATAAT 1862
Db 823 YTCWYTWTHYTHYTHYTHYTHYTHYTHYTHYTHYTHYTHYTHYTHYTHYTHYTHYTHY 868

RESULT 8
CNS0100X CNS0100X 1101 bp DNA linear GSS 26-JUL-1999
LOCUS Drosophila melanogaster genome survey sequence SP6 end of BAC
DEFINITION BACN03G04 of DrosBAC library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL098379
VERSION AL098379.1 GI:5609990
KEYWORDS GSS.
SOURCE Drosophila melanogaster.
ORGANISM Drosophila melanogaster.
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS melanogaster BAC library was prepared by Alain Bucheton
TITLE Direct Submission
JOURNAL Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
COMMENT - web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the European Drosophila Genome Project (EDGP) -
http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
library (Dros BAC) was made by Alain Billaud at CEPH (Centre
d'Etude du Polymorphisme Humain) with funding provided by a MRC
project grant. The DNA was prepared from embryos by Alain Bucheton
and Genevieve Payan. It has been constructed in the vector
pBelOBAC11.
FEATURES
source 1..1101
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="BACN03G04"
/clone_lib="DrosBAC"
/plasmid="pBelOBAC11"
/notes="end : SP6"
BASE COUNT 195 a 108 c 131 g 161 t 506 others
ORIGIN
Query Match 2.5%; Score 47.2; DB 17; Length 1101;
Best Local Similarity 20.9%; Pred. No. 0.36;
Matches 82; Conservative 143; Mismatches 168; Indels 0; Gaps 0;
Qy 519 TTTTATGTCCTCAAGCTTAAAAAATCTGAGAAAAGGATATACCAAGCTTTTGAG 578
Db 325 TTTTNTTNTTBDAGGWATARWAWDCRYATATAAADAACATTTTGTGTDTDKTTT 384
Qy 579 CCATTTTGTATTTCACTTACTGACTAGTATTGGCCTCAGCACTTCTTTTCTCTAGATA 638
```



```

Db 236 GGTAAAGTTACTTGGATGGTATCGTTCCTCCAAAGATCTCGAAATCTCAAAACCTTTAACT 177
Qy 896 GGAATGATTTTAAAGAGGGTATGCGCTTTTACCTTTTAAATCCAAAGAGCAGTGGCATCAGC 955
Db 176 GCAATGCTTGGTGTAAAGTTACCGAGTTTATAGTGTGTAAGGTGGTAAAGTT 117
Qy 956 GATTTGCTCAATGCGTGGCAAGTGTGTCATAACTACAGGAAGTGGTCTTTCTTCT 1015
Db 116 GATTTGCTCAAGGTGGTGTCTTAAATAGATAAATCGATGAAGCTATTCTCGTTCA 57
Qy 1016 AGAGACTTCGTTTCAACATTTGGATCTTCAGAAAGATCAAAATCTA 1065
Db 56 AGAATGCAATGCAACNATTCAAATATCTTATAAATAAATAAATA 7

RESULT 11
CNS00E07 1101 bp DNA linear GSS 04-JUN-1999
LOCUS Drosophila melanogaster genome survey sequence TET3 end of BAC:
DEFINITION BACR29P01 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL069440
VERSION AL069440.1 GI:4949583
KEYWORDS GSS.
SOURCE Drosophila melanogaster.
ORGANISM Drosophila melanogaster.
REFERENCE Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
AUTHORS Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
TITLE Ephydroidea; Drosophilidae; Drosophila.
JOURNAL 1 (bases 1 to 1101)
COMMENT Direct Submission
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see http://www.fruitfly.org The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osoegawa and
Aaron Mammoser in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain y2; cn bw sp, the same strain used for the BDGP's
PI and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES
source
Location/Qualifiers
1..1101
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="BACR29P01"
/clone_lib="RPCI-98"
/notes="end : TET3"
BASE COUNT 366 a 66 c 104 g 351 t 214 others
ORIGIN
Query Match 2.4%; Score 45.4; DB 17; Length 1101;
Best Local Similarity 35.2%; Pred. No. 1;
Matches 107; Conservative 47; Mismatches 150; Indels 0; Gaps 0;

Qy 477 TAGATGCAATGTTCTGTATTCGAAAGAAATCCGAGGAGATTTTATGTCTCAAAAGTC 536
Db 916 WAAATTTTAAWAAWCAATTTTAAWTTAATTTTATTTATACATTTTAAWAAAAATWTT 857
Qy 537 TTAAAAATCTGAGAAAAGGATATACCAAGGCTTTTGGAGCCATTTTTCATCTCACT 596
Db 856 AAATTAWAAAAAARWATTAWAAATTTATATATATTAWAAWTTATATAWTTTAAWAAATW 797

```

```

Qy 597 TGACTACAGTATTGGCCTCAGCACTTCTTTTCTCTAGATACAGGCGCTATTAAAGGT 656
Db 796 TTATAMNATTTTATTTTAAWAAWTTTATWATWATTAATTTTAAATTTTAAATAAAAAA 737
Qy 657 TTGCTTTGACATGATTTTAGGATTTTCTCTCAATGTTTACGGCTCTTTTCATGACTA 716
Db 736 TTTATTTTATTTTATTTTAAWAAWTTTATTTTAAATTTTAAATTTTAAWTTTAA 677
Qy 717 AATTTTCTTCTCATGCTCTGGATGAATAAGCAACCAACATACACAGTTTCATATGATGAATA 776
Db 676 AAWAATTAATTAATAAATTAATTTTAAATTTTAAATTTTAAWTTTAAWTTTATAWTT 617
Qy 777 AGTT 780
Db 616 TTTT 613

RESULT 12
AL566565
LOCUS AL566565 LTI FL013 FBRn1 Homo sapiens cDNA clone CS0DF020YH23 3
DEFINITION prime, mRNA sequence.
ACCESSION AL566565
VERSION AL566565.1 GI:12919043
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
JOURNAL Full-length cDNA libraries and normalization
COMMENT Unpublished (2001)
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
E-mail: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES
source
Location/Qualifiers
1..963
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="LTI FL013 FBRn1"
/dev_stage="pooled tissue from post conception fetuses (20
week_24 week and 26 week)"
/lab_host="DH10B"
/notes="Organ: fetal brain; Vector: pCMVSPORT 6; 1st strand
cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-stranded cDNA was digested with Not I
and cloned into the Not I and Eco RV sites of the
pCMVSPORT 6 vector. Library was constructed by Life
Technologies. Contact : Feng Liang Life Technologies, a
division of Invitrogen 9800 Medical Center Drive Rockville
, Maryland 20850, USA Fax : (1) 301 610 8371 Email :
fliang@lifetech.com URL :
http://fulllength.invitrogen.com"
BASE COUNT 200 a 141 c 94 g 401 t 127 others
ORIGIN
Query Match 2.4%; Score 45.2; DB 9; Length 963;
Best Local Similarity 37.0%; Pred. No. 1.1;
Matches 164; Conservative 59; Mismatches 213; Indels 7; Gaps 2;

Qy 1262 ATCGGGCTTTTAGAGCTTTGGCAATCATCTGCTCATGTGAGTTTGGCTTTGAATGG 1321
Db 88 ATCTTKKATTAATATCATTTTTCYTATACYTAYATTTTTTTTTTTTTTTTTTTTTT 147
Qy 1322 CAATATGCTTTTCAAGTCGGTATGCGCTTTAAATCATGACCTTTTGGCTACTGTGAGTC 1381
Db 148 ATTAAWTTTATTTATTTTATTAATTTTAAATTTKWTCTATTTTATTTTATTTTAT 207
Qy 1382 TTGTTTATAGCACATTTCTTTTGAAGAAAAATTCAAATAGA-----TTTGAAGCCATTG 1436

```

Db 208 YTTTTTTTCAATATTTTTTTTTTAAAMWTTTTTTTTTAAACATTTTTTAAATATWKT 267

QY 1437 GTGCTTAATGACGTGATATGGGGATTCATTAACAATACCTTGATCATTTTGAATCGTA 1496

Db 268 TKTATTTWATYTKTATTKTKTKTTTTTTTTTTTTTAAATTCATTTTTTTTTTTCWYCWWTW 327

QY 1497 TTGCTGAAGATCGCCACGCAACCGGTGTACCCCTATGCATGTTTAAATGAATGCC 1556

Db 328 TTAAVAAATWTTTTTTTTTKTKWKKCTTKTKTTTTTTTTTTTTTTTTTTTTTAAATGKMW 387

QY 1557 TTCAAAAGACCTGACCGCCAGGTAATGACACAGTACATCTATCAGTTTGTGTA 1616

Db 388 KKTMMKKKKTATKACACCTWKKMKWTTTKATTTTAAATKTKTCCTTTTAAATTTWTTT 447

QY 1617 TGCTTTGTTTAAAGCGGCTCTCTGCTTTAATTTGATTTATTAACCATAGGA 1676

Db 448 T-TTTTWTTCMAAATCTTTTCTTTTWTATWTATWTATTTTATTTTAAAGKXAA 505

QY 1677 TTCTTTAGCACTTATCTCT 1699

Db 506 KTKTKTTAKTKACTKATCTKTT 528

RESULT 13

CNS06PBX/C 968 bp DNA linear GSS 05-JUL-2001

LOCUS T7 end of clone AV0AA013B12 of library AV0AA from strain CBS 379 of

DEFINITION Saccharomyces exiguus, genomic survey sequence.

ACCESSION AL409171.1 GI:12176428

VERSION 1

KEYWORDS GSS.

SOURCE Saccharomyces exiguus.

ORGANISM Saccharomyces exiguus.

REFERENCE Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomyces.

AUTHORS Soucier,J.L., Aigle,M., Artiguenave,F., Blandin,G., Bolotin-Fukuhara,M., Bon,E., Broctier,P., Casaregola,S., de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Loirente,B., Malpertuy,A., Neveuglise,C., Ozler-Kalogeropoulos,O., Potier,S., Saurin,W., Tekata,F., Toffano-Nioche,C., Wesolowski-Louvel,M., Wincker,P. and Weissenbach,J.

REFERENCE 1 (bases 1 to 968)

TITLE Genomic exploration of the hemiascomycetous yeasts: 1. A set of yeast species for molecular evolution studies

JOURNAL FEBS Lett. 487 (1), 3-12 (2000)

MEDLINE 20584711

PUBMED 1152876

REFERENCE 2 (bases 1 to 968)

TITLE Bon,E., Neveuglise,C., Lepingle,A., Wincker,P., Artiguenave,F., Galliardin,C. and Casaregola,S.

JOURNAL Genomic exploration of the hemiascomycetous yeasts: 6. Saccharomyces exiguus

MEDLINE 20584716

PUBMED 1152881

REFERENCE 3 (bases 1 to 968)

TITLE Genoscope.

JOURNAL Direct Submission

MEDLINE Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage, 2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail : seget@genoscope.cns.fr - Web : www.genoscope.cns.fr)

COMMENT This GSS is part of a random genomic sequencing program of thirteen yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces exiguus, Saccharomyces servazii, Zygosaccharomyces rouxii, Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila, Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to 5 kb were prepared and both extremities were sequenced. See keywords for description of this sequence and for the sequence of the other extremity of this insert.

FEATURES

LOCATION/Qualifiers

1..968

misc_feature

complement(2..>967)

/note="similar to Saccharomyces cerevisiae ORF YLR087c (1 putative frameshift(s))"

/evidence="not experimental"

BASE COUNT 455 a 148 c 105 g 191 t 69 others

ORIGIN

Query Match 2.4%; Score 45.2; DB 17; Length 968;

Best Local Similarity 38.0%; Pred. No. 1.1;

Matches 146; Conservative 35; Mismatches 203; Indels 0; Gaps 0;

QY 1348 TTTAATTCAGACCTTTGGCTACCTGTCAGCTTGTATAGACATTTCTTTTGA 1407

Db 546 TTTAKGCAATGACATTTTGTCTTCCTGTTACTTAAGTAGATTATCTTTGG 487

QY 1408 GAAATTCAAATAGATTGCAAGCATGGTGTATATGACTGATTTGGGATTCATT 1467

Db 486 WGAATTTGTATCTTKATACATTTGGTGCTTTTAAATTTTCTTTTWTWTTT 427

QY 1468 AAACAATCTTTGATCATTTTGTATGCTATTCGTAAGATCGCCAGCAACTGTTTAC 1527

Db 426 TATTTTGTGKACTTTTGTGTTAAATTTTAAATTTTCTTTTCTTTTWTWTTT 367

QY 1528 CCCATGACGTTTATGATTAATAGCCCTCAAAAGCTGACCGGCAAGTAATGAC 1587

Db 366 TTTTAAGTWTATTTATTTTGTGTTTATTTTAAATGTTTGTGATCTTTTTCG 307

QY 1588 AACAGTCAACACTGATACGTTTGTATAGCTTTGTTAAGCGCGCTCTGCTT 1647

Db 306 KTTTAAATGATTATTCATATTTTCTTTCAATTTTCKTKTKGTCGTGCTTTT 247

QY 1648 TAAATTCATTTATTAAGCAATGAGATTTCTTGAACATTTATGCTCTTATAT 1707

Db 246 TTTCTTTTGTGATTTTATGATGKGAATTTTAAATTTTGTGKTTTWTWTTT 187

QY 1708 TGCACCACTCTGTTGTTTAT 1731

Db 186 TGTATATGTTTATTTTWT 163

RESULT 14

LOCUS BE318484 516 bp mRNA linear EST 21-DEC-2000

DEFINITION NF071B06LF1046 Developing leaf Medicago truncatula cDNA clone

ACCESSION BE318484

VERSION BE318484.2 GI:11959810

KEYWORDS EST.

SOURCE barbel medic.

ORGANISM Medicago truncatula

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicot; Rosidae; eustosids I; Fabales; Fabaceae; Papilionoideae; Trifoliaceae; Medicago.

AUTHORS 1 (bases 1 to 516)

TITLE Flores-Derez,I., Scott,A.D., Harris,A.R., Gonzales,R.A., Bell,C.J., Flores,H.R., Imman,J.T., Weller,J.W. and May,G.D.

JOURNAL Expressed Sequence Tags from the Samuel Roberts Noble Foundation Medicago truncatula leaf library

COMMENT On Jul 14, 2000 this sequence version replaced gi:9192261.

Contact: May GD

Plant Biology Division

The Samuel Roberts Noble Foundation

2510 Sam Noble Parkway, Ardmore, OK 73402, USA

Tel: 580 221 7391

BASE COUNT
ORIGIN

415 a 142 c 148 g 345 t 75 others
http://arrington.invercogen.com

Query Match 2.3%; Score 43.8; DB 9; Length 1125;
 Best Local Similarity 45.0%; Pred. No. 2.8;
 Matches 108; Conservative 11; Mismatches 132; Indels 0; Gaps 0;

QY 484 AAATGCTTGTATTCGAAAGATCCGAGAGAAATTTATGTCATGACCTTAATAA 543
 DB 891 AAATTTTAAATATATTAATAAATAAATAAATAAATAAATAAATAAATAA 832
 QY 544 ATCTGAGAAAAGATATACCAAGCTTTGAGCCATTTTGAATCTGACTGAC 603
 DB 831 TWTATATATAAATAATATATAAATAAATAAATAAATAAATAAATAAATAA 772
 QY 604 AGATATGCGCTCAGACCTCTTTCTTCTGATACAGGCGCTATTAAGGTTGCTT 663
 DB 771 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 712
 QY 664 GACATGATTTTAGAATTTCTCTCAATGTTAGCGCTCTTTTATGACATAATTTT 723
 DB 711 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 652
 QY 724 CTTCATGCTGT 734
 DB 651 TTTTATTTTATTTT 641

RESULT 17
 BQ399330 417 bp mRNA linear EST 22-MAY-2002
 LOCUS BQ399330
 DEFINITION NISC mp03a03.x1 NICHD XGC Spl Xenopus laevis cDNA clone
 IMAGE:4965412 3', mRNA sequence.
 ACCESSION BQ399330
 VERSION BQ399330.1 GI:21087017
 KEYWORDS EST.
 SOURCE African clawed frog.
 ORGANISM Xenopus laevis
 Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae; Pipridae;
 Xenopodinae; Xenopus.
 1 (bases 1 to 417)
 REFERENCE NIH-XGC http://image.llnl.gov/image/html/xenopuslib.info.shtml.
 AUTHORS National Institute of Child Health and Human Development, National
 Cancer Institute, Xenopus Gene Collection
 TITLE Unpublished (2002)
 JOURNAL Contact: Robert Strausberg, Ph.D.
 COMMENT Email: cgapbs-remail.nih.gov
 cDNA Library Preparation:
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium/LLNL
 DNA Sequencing by: National Institutes of Health Intramural
 Sequencing Center (NISC)
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 info@image.llnl.gov
 Plate: LLAM10941 row: B column: 5
 Seq primer: -21M13 forward primer (ABI).
 Location/Qualifiers
 1. 417

FEATURES
 source
 1. 417
 /organism="Xenopus laevis"
 /db_xref="taxon:8355"
 /clone="IMAGE:4965412"
 /clone_lib="NICHD XGC Spl"
 /dev_stage="adult"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: spleen; Vector: PCMV-SORT6; Site 1: NotI;
 Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt.
 Average insert size 2.4 kb. Constructed by Life
 Technologies."

BASE COUNT 174 a 56 c 2 g 185 t

ORIGIN

Query Match 2.3%; Score 43.6; DB 14; Length 417;
 Best Local Similarity 49.1%; Pred. No. 2.2;
 Matches 115; Conservative 0; Mismatches 119; Indels 0; Gaps 0;

QY 1608 TTTTGTATATGCTTTGTTATATAGCGGCTCTGCTTATTTGCAATTATATGA 1667
 DB 31 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 90
 QY 1668 CCATAGGATTCCTTCGAAATCTTATGCTCTGTTATATGACGACCTCTGTTGT 1727
 DB 91 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 150
 QY 1728 TTATGTCGCTAAAGAAATCGCTCAAAATAGTACCGTTAACTTAATCTAACGTTAG 1787
 DB 151 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 210
 QY 1788 CAATATTAATATCTCTTGGGACTTATGCTCCAAAGGCCCTGTGTATTA 1841
 DB 211 AAAAAAAAAATATAATTTTAACTTTACCCGCCGCCCTTCTTAATAA 264

RESULT 18
 AU087215 500 bp mRNA linear EST 27-JAN-2001
 LOCUS AU087215
 DEFINITION AU087215 Sugano Malaria cDNA library Plasmodium falciparum 3D7 cDNA
 clone XPFn4019, mRNA sequence.
 ACCESSION AU087215
 VERSION AU087215.1 GI:12389356
 KEYWORDS EST.
 SOURCE Plasmodium falciparum 3D7.
 ORGANISM Plasmodium falciparum 3D7
 Eukaryote; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 1 (bases 1 to 500)
 REFERENCE Watanabe,J., Sasaki,M., Suzuki,Y. and Sugano,S.
 FUL-malaria: a database for a full-length enriched cDNA library
 from human malaria parasite, Plasmodium falciparum
 Nucleic Acids Res. 29 (1), 70-71 (2001)
 20574754
 JOURNAL Contact: Junichi Watanabe
 MEDLINE Institute of Medical Science
 COMMENT The University of Tokyo, Department of Parasitology
 4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
 Tel: 81-3-5449-5378
 Fax: 81-3-5449-5410
 Email: jwatanabe@med.s.u-tokyo.ac.jp
 Suzuki,Y., Yoshitomo-Nakagawa,K., Maruyama,K., Suyama,A. and Sugano
 S. Construction and characterization of a full length-enriched and
 a 5'-end-enriched cDNA library Gene 200 (1-2), 149-156 (1997).
 Location/Qualifiers
 1. 500
 /organism="Plasmodium falciparum 3D7"
 /strain="3D7"
 /db_xref="taxon:36329"
 /clone="XPFn4019"
 /clone_lib="Sugano Malaria cDNA library"
 /dev_stage="erythrocytic stage"

BASE COUNT 172 a 27 c 23 g 274 t 4 others

Query Match 2.3%; Score 43.6; DB 9; Length 500;
 Best Local Similarity 47.7%; Pred. No. 2.3;
 Matches 124; Conservative 0; Mismatches 136; Indels 0; Gaps 0;

QY 484 AAATGCTTGTATTCGAAAGATCCGAGAGAAATTTATGTCATGACCTTAATAA 543
 DB 208 AAATTTTAAATATATTAATAAATAAATAAATAAATAAATAAATAAATAA 267
 QY 544 ATCTGAGAAAAGATATACCAAGCTTTGAGCCATTTTGAATCTGACTGACTAC 603
 DB 268 TTTTAAATTAATATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 327
 QY 604 AGATATGCGCTCAGACCTCTTTCTTCTGATACAGGCGCTATTAAGGTTGCTT 663
 DB 328 ATTTTATGATGATTTATATCTTTCTTTTAAATACGAAATCGCTTTGTTTCTT 387
 QY 664 GACATGATTTTAGAATTTTCTCTCAATGTTTAGCGCTCTTTTATGACTAAATTTT 723

OY	1405	GAGAAGAAATTCAGTAAGATTGCGACGGCTTGAATGAGCTGTATGCGGATTC	1464
Db	726	TKASHSTBSCAACGAMATTTTWTMTKMMCAACAACKKKMMAAAAAAABAAAAXXXXXXTM	785
OY	1465	ATTAAACATACTTGATTCATTTTGTGATCGTATGCGTGAAGATCGCAAGCAACTGTT	1524
Db	786	AATMACATADTADTATTATTTKTITTAABKBKRBDNMAAMCCCCAKTKTATTTTT	845
OY	1525	TACCCC	1530
Db	846	TCSGCC	851
 RESULT 21 CNS00DM2/c			
LOCUS	CNS00DM2	803 bp	DNA linear GSS 04-JUN-1999
DEFINITION	Drosophila melanogaster genome survey sequence TET3 end of BAC # BACR27016 of RPc1-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.		
ACCESSION	AL072713		
VERSION	AL072713.1	GI:4948210	
KEYWORDS	GSS.		
SOURCE	Drosophila melanogaster.		
ORGANISM	Drosophila melanogaster. Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.		
REFERENCE	Genoscope. Direct Submissio		
AUTHORS	Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage		
TITLE	Bp 191 91006 EVER cedex - FRANCE (E-mail : secretfgenoscope.cns.fr		
JOURNAL	- Web : www.genoscope.cns.fr) Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazuo Osoegawa and Aaron Mammosser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPc1-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2; cn bw sp, the same strain used for the BDGP's p1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPac Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.		
 FEATURES			
source	location/Qualifiers		
	1..803	/organism="Drosophila melanogaster"	
	/db_xref="taxon:7227"		
	/clone="BACR27016"		
	/clone_id="RPc1-98"		
	/note="end : TET3"		
BASE COUNT	248 a 94 c 131 g 158 t	172 others	
ORIGIN			
Query Match	2.3%; Score 43.2; DB 17; Length 803;		
Best Local Similarity	31.9%; Pred. No. 3.5;		
Matches	150; Conservative 60; Mismatches 260; Indels 0; Gaps 0;		
OY	1032	AAACATTGGATCTTCAGAAAAGATCAAAATCTATTATTAGATAAAGCTTTAAGCTATA	1091
Db	663	RAAHHTTWTAAHATTHMRSGBATFAAAAAWWTTTAAAWWTAAMAGKTTGCGRCRTAM	604
OY	1092	CTAAGCAGATCAGAGCTCTCTCTCTAAATTAACATCATGAGCTGGCTTATGTGGGA	1151
Db	603	AAAAAAAAAACRCGGHCATATAAAAAHAHTCAKTTCCCMCGTCKTMTAGAGWKKGXTR	544
OY	1152	TGTGTTGTGAGAAACAGGCCTAGATTCTCTCAGCGAAGACTCTAAGCAACGCAAAATTT	1211
Db	543	KTKCAAAGGGGGGKGTCCGCHYTARHRGRKWAAGTCTCAATTTTAAAGAAARATTT	484

```

Oy 1212 GGTCAAAAGGTAGAGCAAACTATCGAAGAAAATGCGTTACAGCGCACTCGGGCTTT 1271
      :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 483 CRTAGTGAAAAAARGCAARAAAAAAMAKAGBMTDDTTTGTAKTARGRARAC 424
      :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Oy 1272 TAGGAGCTTTGGCAATCATCTTGTCTATGTGAAGTTGGCGCTTGAATGGCAATATGCTT 1331
      :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 423 AATACCCKTMVGMVTTTTTGGYKTTTWTWKYCKTKKCCACACNNCGNNATCGTGCTT 364
      :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Oy 1332 TCAGTGCCTATGAGCGCTTTAATTCATGACCTTTTGGCTACTGTGCAGCTCTGTATTAG 1391
      :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 363 TCTCTGAGAGTCTTTTATGTCTATGTCTTACACAMTTTTTNNNNNNNNCTTTTATT 304
      :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Oy 1452 TATTGGGCTATCTTAACAATCTTGTATCATCTTTTGTATGCTATTGCT 1501
      :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 243 TATCTCTCANTATATATAAAAAACCACTCATCTATCTTTCCNNTACAT 194
      :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

RESULT 22
LOCUS A2550729 863 bp DNA linear GSS 14-NOV-2000
DEFINITION Entamoeba histolytica Sheared DNA Entamoeba histolytica
ACCESSION A2550729
VERSION A2550729 GI:11176030
KEYWORDS GSS.
SOURCE Entamoeba histolytica.
ORGANISM Entamoeba histolytica.
REFERENCE 1 Entamoeba histolytica.
AUTHORS Eukaryota; Entamoebidae; Entamoeba.
TITLE 1 (Bases 1 to 863)
COMMENT Loftus,B., Van Aken,S. and Fraser,C.
JOURNAL Determination of clone end sequences from Entamoeba histolytica
COMMENT HML:IMSS sheared DNA library
COMMENT Unpublished (2000)
COMMENT Contact: Brendan J Loftus
COMMENT Department of Eukaryotic Genomics
COMMENT The Institute for Genomic Research
COMMENT 9712 Medical Center Dr., Rockville, MD 20850, USA
COMMENT Tel: 301 838 0208
COMMENT Fax: 301 838 3543
COMMENT Email: B.loftus@icigr.org
COMMENT Clones are derived from the Entamoeba histolytica HML:IMSS sheared
COMMENT DNA library
COMMENT Seq primer: M13-Reverse
COMMENT Class: shotgun
COMMENT High quality sequence start: 20
COMMENT High quality sequence stop: 794.
FEATURES
    source
        1..863
            /organism="Entamoeba histolytica"
            /strain="HML:IMSS"
            /db_xref="taxon:5759"
            /clone_lib="Entamoeba histolytica Sheared DNA"
            /note="Vector: pHOSt; Site: 1; Bst I; Constructed at The
            Institute for Genomic Research (TIGR), Rockville, MD.
            Genomic DNA isolated from broth cultures of E. histolytica
            using a method described by Clark and Diamond (Clark,
            C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
            method for isolate identification. Exp. Parasitol.
            77(450).). The DNA was mechanically sheared to give a
            tight size distribution (~2 Kd). The v + i method used for
            the library construction is described in detail in Smith,
            H.O. and Venter, J.C. (Making small insert libraries for
            whole genome shotgun sequencing projects. In Genome
            Sequencing: A Practical Approach, eds M. Vaudin and B.
            Borell, Oxford University Press, 1999)."
```

```
Query Match      2.3%; Score 43.2; DB 17; Length 863;
Best Local Similarity 47.1%; Pred. No. 3.6;
Matches 132; Conservative 0; Mismatches 148; Indels 0; Gaps 0;

QY 523 ATTGCTCAAGTCTTAAATAATCTGTAGAAAAGGATATACCAAGGCTTTGGAGCCAT 582
DB 334 ATGAATTAATCTATTATGATATTAGAGAAATGATATCAAGTAATATAGTATA 393
QY 583 TTTTGATCTTAACCTGACTACAGTATTGGCTCAGCACTCTCTTTCTCTTCTAGATACAGG 642
DB 394 TTACAAATATATGATGAATAAATATTATTAAATAAATACATTTTCATCAATGTATCATT 453
QY 643 GCCTATTAAAGGTTTGGCTTTGACATTTAGTATTTAGGAAATTTCTCTTCAATGTTTACGGC 702
DB 454 TTGTAGTATTGATATTGATTTATTTCAATTTATGTAATATATACATTTATTATTATTC 513
QY 703 TCTTTTCATGACTAAATTTTCTTCATGCTGTGATGAATAAGCAACCAACATACACAGTT 762
DB 514 TTTTCATCTTCATTAATATCAATTTAAATAGAAATATTAATCAATATGAAACAATATA 573
QY 763 GCATATGATGAATAAGTTCGTGGGATAAAGCATGATTTTC 802
DB 574 AATAATGACAAATGATTATATACTAAATAAGAAATATTTTC 613

RESULT 23
BGJ50812/c
LOCUS
DEFINITION      332 bp mRNA linear EST 01-MAR-2001
sequence.
ACCESSION      BGJ50812
VERSION        BGJ50812.1 GI:13179554
KEYWORDS
SOURCE
ORGANISM        Solanum tuberosum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
REFERENCE
1 (bases 1 to 332)
NIELSEN,K.L, Crookshanks,M., Emmersen,J. and Welinder,K.G.
EST-sequencing of mature potato tuber (Var. Kuras)
Unpublished (2000)
Contact: Karen G. Welinder
Institut for bioteknologi
Aalborg Universitet
Sohngaardsholmsvej 49, 9000 Aalborg, Denmark
Tel: +45 96358467
Fax: +45 98141808
Email: kgw@bio.auc.dk
Sequenced from the 5' end.
High quality sequence stop: 332
POLYA=No.
FEATURES
Location/Qualifiers
source
1..332
/organism="Solanum tuberosum"
/cultivar="Field grown Kuras"
/db_xref="taxon:4113"
/clone_lib="Mature tuber lambda ZAP"
/tissue_type="Tuber"
/notes="Vector: Lambda ZAP"
BASE COUNT      69 a 60 c 104 g 99 t
ORIGIN
Query Match      2.3%; Score 42.8; DB 12; Length 332;
Best Local Similarity 56.3%; Pred. No. 3.3;
Matches 80; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 1457 GGGTATTCAATTAACAATCTTTGATCATCTTTTGTATCGTATTCGTGAAGATCGCAAGCG 1516
DB 146 GAGCAGTATTCAACAGCGGCTTCATCCCTTTAGCTCCAAAAGCACAAGATCCACAGAT 87
QY 1517 AACCTGTTACCCCTATGTCATGTTTGTATTAATGATGCCCTTCAAAGACGTTTCAGCGCG 1576
```

```
DB 86 AGCTTCTTCAACCCCTTACAAATCTTCGCCAAAAGATTCATTCAGTATCAGTAAGCTCC 27
QY 1577 ACGTAAATGCAACAGCTACAA 1598
DB 26 CGGSCCATGACCGGAGCTTCAA 5

RESULT 24
CNS0155H
LOCUS
DEFINITION      1001 bp DNA linear GSS 26-JUL-1999
Drosophila melanogaster genome survey sequence SP6 end of BAC
BACN13C23 of DrosBAC library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION      AL105023
VERSION        AL105023.1 GI:5617037
KEYWORDS
SOURCE
ORGANISM        Drosophila melanogaster.
Drosophila melanogaster.
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
1 (bases 1 to 1001)
Genoscope.
Direct Submission
Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the European Drosophila Genome Project (EDGP) -
http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
library (Dros BAC) was made by Alain Billaut at CEPH (Centre
d'Etude du Polymorphisme Humain) with funding provided by a MRC
project grant. The DNA was prepared from embryos by Alain Bucheton
and Genevieve Payan. It has been constructed in the vector
pBelobAC11
FEATURES
Location/Qualifiers
source
1..1001
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="BACN13C23"
/clone_lib="DrosBAC"
/plasmid="pBelobAC11"
/notes="end : SP6"
BASE COUNT      266 a 219 c 134 g 150 t 232 others
ORIGIN
Query Match      2.3%; Score 42.6; DB 17; Length 1001;
Best Local Similarity 29.7%; Pred. No. 5.5;
Matches 126; Conservative 85; Mismatches 213; Indels 0; Gaps 0;

QY 1348 TTAAATTCATGACCTTTTGGCTACCTGTCAGTCTTCTTTATAGCACATTTCTTTTGA 1407
DB 911 WWAATTTTAAATTTTATTTTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTAA 852
QY 1408 GAAATTCAAATAGATTGCAAGCCATTGGTGCTTTTAAATGACTGTATTTGGGGTATTCATT 1467
DB 851 AAAATTTAAATTTTWTWTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTT 792
QY 1468 AACAATACCTTTGATCATTTTGTATTCGTATTCGTGAAGATCGCAAGCAACCTGTTTAC 1527
DB 791 TAAWTTTTTTTATTTTAAATTTTATTTTAAATTTTAAATTTTAAATTTTAAATTTT 732
QY 1528 CCCTATGCATGTTTATGATGATGCCCTTCAAAGACGTTTCAGCCGACGGTAATGAC 1587
DB 731 HYATTTTATTTATTTATTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 672
QY 1588 AACAGCTACAACTCTATCAGTTTGTGTTTAAATGCTTTTGTATAGGCGGCTCCTCTGCTT 1647
DB 671 MCYTTTWTDBNTTTTGTGTTTWTHTTWTHTTWTHTTWTHTTWTHTTWTHTTWTHTT 612
QY 1648 TAATTTGCAATTTATATGACCATAGGATTTCTCTAGGAACATTTATCGTCTCTTTATAT 1707
DB 611 MAMYTTTSTYCTHTVTTTTCYCMTTTTTCTCTNTATTTCTTCTTCTTCTTCTTCTTCTTCTT 552
```



```

Db 202 TTAGACACKTGGCAATTAATTTTGGTKTKTGGTTATATWTTWTKGTAAMGKTAWTTTT 143
Qy 1817 TCC 1819
Db 142 TTC 140

RESULT 31
CNS04E5M          976 bp  DNA  linear  GSS 21-MAY-2000
LOCUS             Tetraodon nigroviridis genome survey sequence T7 end of clone
DEFINITION        103P02 of library G from Tetraodon nigroviridis, genomic survey
sequence.
ACCESSION         AL286627
VERSION           AL286627.1 GI:8025084
KEYWORDS          GSS; genome survey sequence.
SOURCE            Tetraodon nigroviridis.
ORGANISM          Tetraodon nigroviridis
                  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                  Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
                  Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
                  Tetraodontidae; Tetraodon.
REFERENCE         1 (bases 1 to 976)
AUTHORS           Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
                  Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
                  Saurin,W. and Weissenbach,J.
TITLE             Human gene number estimate provided by genome wide analysis using
                  Tetraodon nigroviridis DNA sequence
JOURNAL           Unpublished
REFERENCE         2 (bases 1 to 976)
AUTHORS           Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
                  Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
                  Weissenbach,J.
TITLE             Characterization and repeat analysis of the compact genome of the
                  freshwater pufferfish Tetraodon nigroviridis
JOURNAL           Unpublished
REFERENCE         3 (bases 1 to 976)
AUTHORS           Genoscope.
TITLE             Direct Submission
JOURNAL           Submitted (12-APR-2000)
COMMENT           This sequence is a single read and was generated as part of a large
                  scale clone-end sequencing project of the Tetraodon nigroviridis
                  genome. For more information, please take a look at
                  http://www.genoscope.cns.fr/Tetraodon.
FEATURES          Location/Qualifiers
source            1..976
                  /organism="Tetraodon nigroviridis"
                  /db_xref="taxon:99883"
                  /clone="103P02"
                  /clone_lib="G"
                  /notes="Genoscope sequence ID : COBG103DH01LP1-end : T7"
BASE COUNT       243 a 151 c 137 g 327 t 118 others
ORIGIN

Query Match      2.2%; Score 41.8; DB 17; Length 976;
Best Local Similarity 38.2%; Pred. No. 8.8;
Matches 109; Conservative 28; Mismatches 148; Indels 0; Gaps 0;

Qy 517 ATTTTATTGCTCAAGCTCTTAAATAATCTGTAGAAAAGCATATACCAAGGCTTTGG 576
Db 678 ATTTATTWTTTAAATNNAWATTAATAAATTTTAWTTWATATAATTAATTAATTA 737
Qy 577 AGCCATTTTGATCTTAACCTGTACAGTATGCGCTCAGCAGCTCTTTCTTCCTCCTAGA 636
Db 738 WATWTTWAWATTTTWTWAWATTTWTTWATNATATWTTWATTAATTTTATTA 797
Qy 637 TACAGGCGCTATTAAAGGGTTTGGCTTTGACATTTAGGAATTTTCTCTTCAATGTT 696
Db 798 TTAATAWATTAATAAATAAATWATTTTTTTTATAAATTTTATTAATWAWWTTWAWATTT 857
Qy 697 TACGCGCTTTTCAGTAAATTTTCTTCATGCTGTGGATGAATAGACCCCAACATAC 756

```

```

Db 858 TATTATTWTTWTTTANAANAATATATATAATWATTTTNTWAAATTTAATTTTAAATTA 917
Qy 757 ACGTTGCATATGATGAATAGTTTCGTGGGATAAAGCATGATTT 801
Db 918 TAAATATAATATTATTTTATTTTATTAATAAATTTAAATAATTTT 962

RESULT 32
CNS051P2          1096 bp  DNA  linear  GSS 26-MAY-2000
LOCUS             Tetraodon nigroviridis genome survey sequence T7 end of clone
DEFINITION        024H02 of library A from Tetraodon nigroviridis, genomic survey
sequence.
ACCESSION         AL339167
VERSION           AL339167.1 GI:8232925
KEYWORDS          GSS; genome survey sequence.
SOURCE            Tetraodon nigroviridis.
ORGANISM          Tetraodon nigroviridis
                  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                  Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
                  Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
                  Tetraodontidae; Tetraodon.
REFERENCE         1 (bases 1 to 1096)
AUTHORS           Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
                  Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
                  Saurin,W. and Weissenbach,J.
TITLE             Human gene number estimate provided by genome wide analysis using
                  Tetraodon nigroviridis DNA sequence
JOURNAL           Unpublished
REFERENCE         2 (bases 1 to 1096)
AUTHORS           Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
                  Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
                  Weissenbach,J.
TITLE             Characterization and repeat analysis of the compact genome of the
                  freshwater pufferfish Tetraodon nigroviridis
JOURNAL           Unpublished
REFERENCE         3 (bases 1 to 1096)
AUTHORS           Genoscope.
TITLE             Direct Submission
JOURNAL           Submitted (12-APR-2000)
COMMENT           This sequence is a single read and was generated as part of a large
                  scale clone-end sequencing project of the Tetraodon nigroviridis
                  genome. For more information, please take a look at
                  http://www.genoscope.cns.fr/Tetraodon.
FEATURES          Location/Qualifiers
source            1..1096
                  /organism="Tetraodon nigroviridis"
                  /db_xref="taxon:99883"
                  /clone="024H02"
                  /clone_lib="A"
                  /notes="Genoscope sequence ID : COAA024D01C1-end : T7"
BASE COUNT       488 a 127 c 179 g 236 t 66 others
ORIGIN

Query Match      2.2%; Score 41.8; DB 17; Length 1096;
Best Local Similarity 40.3%; Pred. No. 9.1;
Matches 127; Conservative 0; Mismatches 188; Indels 0; Gaps 0;

Qy 1420 AGATTTCGAAGCCATTGGCTGCTTTAATGACTGTATTCGGGTATTCATTAACAATACTTT 1479
Db 455 AGAAGTGGCTGCGTNNNGTACTCACTTTCCACCACATTTGTATCTTTCAACAGAAATCAT 396
Qy 1480 GATCATTTTTCATCGTATTCGTGAAGATCCCAAGCAACCTGTTTACCCCTATGCATGT 1539
Db 395 NATWTGCTTTTNTTTTATNCNTNTNCAGCTTAAGCAACCCANTTTTNCNNGAAACNTTTT 336
Qy 1540 TTTAGTTAATGATGCCCTTCAAAGAGCTTCAGCGCACGGTAAATGACACAGCTACAAC 1599
Db 335 TTTTACTCCNTNCAAGANNNNNNNTTTTTTTTTNTNTNTTTTTTTTTCNNTNTTTTCAAG 276
Qy 1600 TCTATCAGTTTGTGTAATGCTTTGTTTATAGGGCGGCTCCTGCTCTTTAATTTGCAAT 1659
Db 275 TTNNTTTTTTTTTTTTTNTNTNTNTTTTTTTTTTTTTTTTTTTTTNTNTNTNTTTTTTTT 216

```


QY 523 ATTGCTCAAGTCTTAAAAATCTGTAGAAAAAGGATATACCAAGGCTTTTGGAGCCAT 582
 Db 210 ATGGAATTAATCTATATATGATATTTAGAGAGAAATGAATAATCAATATATATAGTATA 269
 QY 583 TTTTGATTCTAACTTGACTACAGTATTCGCTCAGCACTCTTTTCTTCTCCCTAGATACAGG 642
 Db 270 TTACAAATATATGATGAATAAATATATTAATATATAAATACATTTTTCATCAATGTATCAAT 329
 QY 643 GCCTATTAAAGGTTTGGCTTTGACATTTAGCAATTTTCTCTCAATGTTTACGGC 702
 Db 330 TTGTAGTATTGATATTGATATTCTTCAATTAATGTAATATACATTTATTTATTC 389
 QY 703 TCTTTTCATGACTAAATTTTCTTCATGCTGTGGATGAATAGAACCCCAACATACAGATT 762
 Db 390 TTTCACTCTTCATTAATATCAATTTTAAATAGAAATATAATTAACAATAATGAACAATATA 449
 QY 763 GCATATGATGAATAGTTTCGTGGGATAAAGCATGATTTTC 802
 Db 450 AATAATGCAAAATGATTAATACTAAAATGAATATTTTC 489

RESULT 35
 AZ541040
 LOCUS ENTDD56TR Entamoeba histolytica Sheared DNA linear GSS 14-NOV-2000
 DEFINITION genomic, DNA sequence.
 ACCESSION AZ541040
 VERSION AZ541040.1 GI:11148393
 KEYWORDS GSS.
 SOURCE Entamoeba histolytica.
 ORGANISM Entamoeba histolytica.
 Eukaryota; Entamoebidae; Entamoeba.
 Loftus, B., Van Aken, S. and Fraser, C.
 REFERENCE 1 (bases 1 to 931)
 AUTHORS Determination of clone end sequences from Entamoeba histolytica
 TITLE HM1:INSS sheared DNA library
 JOURNAL Unpublished (2000)
 COMMENT Contact: Brendan J Loftus
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0208
 Fax: 301 838 3543
 Email: bjlloftus@tigr.org
 Clones are derived from the Entamoeba histolytica HM1:INSS sheared
 DNA library
 Seq primer: M13-Reverse
 Class: shotgun
 High quality sequence start: 89
 High quality sequence stop: 909.
 Location/Qualifiers
 1..931
 /organism="Entamoeba histolytica"
 /strain="HM1:INSS"
 /db_xref="taxon:5759"
 /clone_lib="Entamoeba histolytica Sheared DNA"
 /note="Vector: pHS01; Site 1: Bst I; Constructed at The
 Institute for Genomic Research (TIGR), Rockville, MD
 Genomic DNA isolated from broth cultures of E. histolytica
 using a method described by Clark and Diamond (Clark,
 C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
 method for isolate identification. Exp. Parasitol.
 77:450.). The DNA was mechanically sheared to give a
 tight size distribution (~2 kb). The v + i method used for
 the library construction is described in detail in Smith,
 H.O. and Venter, J.C. (Making small insert libraries for
 whole genome shotgun sequencing projects. In Genome
 Sequencing: A Practical Approach, eds. M. Vaudin and B.
 Barrell, Oxford University Press, 1999)."
 262 a 174 c 99 g 396 t

FEATURES
 source
 1..931
 /organism="Entamoeba histolytica"
 /strain="HM1:INSS"
 /db_xref="taxon:5759"
 /clone_lib="Entamoeba histolytica Sheared DNA"
 /note="Vector: pHS01; Site 1: Bst I; Constructed at The
 Institute for Genomic Research (TIGR), Rockville, MD
 Genomic DNA isolated from broth cultures of E. histolytica
 using a method described by Clark and Diamond (Clark,
 C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
 method for isolate identification. Exp. Parasitol.
 77:450.). The DNA was mechanically sheared to give a
 tight size distribution (~2 kb). The v + i method used for
 the library construction is described in detail in Smith,
 H.O. and Venter, J.C. (Making small insert libraries for
 whole genome shotgun sequencing projects. In Genome
 Sequencing: A Practical Approach, eds. M. Vaudin and B.
 Barrell, Oxford University Press, 1999)."
 262 a 174 c 99 g 396 t

BASE COUNT
 ORIGIN

Query Match 2.2%; Score 41.6; DB 17; Length 931;
 Best Local Similarity 53.3%; Pred. No. 9.7;
 Matches 112; Conservative 0; Mismatches 94; Indels 4; Gaps 1;
 QY 1586 ACAACAGCTCAACCTCATCAGTTTCTTAATGCTTTTGTATTATAGCGGCTCTCTGTC 1645
 Db 127 ACAAGAGACATACTCAATATATTTTAAAGGCTAATTTTTTGTCTTATCAATCTC 186
 QY 1646 TTTAATTTTGCAATTTATGACCATAGGATTTCTTAGGAATCTTTATCGTCTCTTTAT 1705
 Db 187 CTTAAT---CTTTTTCATATACACAAATTAATTAATTCCTTTTCTGTTATTAAT 242
 QY 1706 ATTGCACACCTCTGTTGTTTATGTCGCTAAAGAAATCGCTCAAAATAGTACCG 1765
 Db 243 TTATAAAAATCTCCCTTTTGTGTTTTTAAACAATTAATAAATCTAAACAAACAAATACAA 302
 QY 1766 TTAAACTTAATCTAACGTGTAGCAATATA 1795
 Db 303 TTAATAAATATCATTTGTTTACCAGAAATGA 332

RESULT 36
 CNS00EPP
 LOCUS Drosophila melanogaster genome survey sequence TET3 end of BAC:
 DEFINITION BACR29N07 of RPCI-98 library from Drosophila melanogaster (fruit
 fly), genomic survey sequence.
 ACCESSION AL069494.1 GI:4949637
 VERSION AL069494
 KEYWORDS GSS.
 SOURCE Drosophila melanogaster.
 ORGANISM Drosophila melanogaster.
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 Ephydroidea; Drosophilidae; Drosophila.
 REFERENCE 1 (bases 1 to 1080)
 AUTHORS Genoscope.
 TITLE Direct Submission
 JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequenage :
 BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
 - Web : www.genoscope.cns.fr)
 COMMENT Determination of this BAC-end sequence was carried out as part of a
 collaboration with the Berkeley Drosophila Genome Project (BDGP).
 The BDGP is constructing a physical map of the Drosophila
 melanogaster genome using these BACs. For further information
 please see <http://www.fruitfly.org/TheBDGP/Drosophila>
 melanogaster BAC library was prepared by Kazutoyo Osoegawa and
 Aaron Mammoler in Pieter de Jong's laboratory in the Department of
 Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
 NY. The library is named RPCI-98 and was constructed by partial
 EcoRI digestion of Drosophila DNA provided by the BDGP from the
 isogenic strain Y2; cn bw sp, the same strain used for the BDGP's
 PI and EST libraries. A more detailed description of the library
 and how to order individual BAC clones, the entire library, or
 filters for hybridization from the BACPAC Resource Center can be
 found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES
 source
 1..1080
 /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"
 /clone_lib="BACR29N07"
 /clone_lib="RPCI-98"
 /notes="end : TET3"
 BASE COUNT 171 a 105 c 219 g 225 t 360 others
 ORIGIN

Query Match 2.2%; Score 41.6; DB 17; Length 1080;
 Best Local Similarity 21.6%; Pred. No. 10;
 Matches 109; Conservative 151; Mismatches 244; Indels 0; Gaps 0;
 QY 417 TGGATGCCCACTCACCTTGTCTAGGACTCGCTGGGATGTTCTTGTCTATGGGATGCCG 476
 Db 453 TGCNNNNNNNNNNCCCTBKGGKGGTTCGNGGGGGGGGGGGTNGKGGKKGKG 512

QY 477 TAGATGCAATGCTCTTCTGATTCGAAAGATCCGAGAGAAATTTTATGTCCTAAAGTC 536
 Db 513 TTKTKKGGGCTTWTWTWTTGGGGGKGATGADATGGTGTGTTTTTTTTTTTCTTDTDACT 572
 QY 537 TTAATAAATCTGTAGAAAAGATATACCAAGGCTTTGGAGCCATTTTGTATCTAACT 596
 Db 573 WKANDRKTWTWTWKKTGTGAGRWMAAGADWTRTDGAAGATGKMWAGRKMDKKK 632
 QY 597 TGACTACAGATGGCCCTCAGCACTCTCTTCTCTAGATACAGGGCCTATTAAGGGT 656
 Db 633 TKKTKWGTBKWTKRKSGDTRKTGTCTKTKDGRKKGKGRDPAAGAKTRKTRKCDT 692
 QY 657 TTGCTTGACATGATTTTATGAAATTTCTCTCAATTTTACGGCTCTTTCACTAGCTA 716
 Db 693 KRRDRAGGGGKGDTTDDKGDGMWADATDWTKKDDTWTWMDATDTRKMWMDRWAKD 752
 QY 717 AATTTTCTTCATGCTGATGATGATAGCAACCAACATACACAGTTCATGATGATA 776
 Db 753 DDDWAGAGAAAGARADRTDADARGMAARAKDKTTAATWTRKTRTDITDGDGRK 812
 QY 777 AGTTCGGGGATTAAGCAGATTTCTTGAGAGATGCAAAAACCTTGGGCTGTTCTG 836
 Db 813 RKGWGRKKGAGAGTRKKGARKGATDAAAAMWMDTRTADKTADFKTTTTTATTTKTK 872
 QY 837 GAATGCTTTTCTTTAGTTCGCTGCTCGGCTTGGAGCCTCGAATTCGTTTGG 896
 Db 873 GGAATGCTGATKATADATATKRWMDKRRDRAARTAMAKGMAKADGKRRRTAGDR 932
 QY 897 GAATGATTTTAAAGAGAGATG 920
 Db 933 GAAATRRMGADTRDWTGKGR 956

RESULT 37
 BF279297 552 bp mRNA linear EST 07-MAR-2001
 LOCUS
 DEFINITION GA_Eb0038F20f Gossypium arboreum 7-10 dpa fiber library Gossypium
 accession BF279297
 version BF279297.1 GI:11210367
 keywords
 source
 organism
 Gossypium arboreum.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids II; Malvales; Malvaceae; Gossypium.
 1 (bases 1 to 532)
 REFERENCE
 AUTHORS Wing, R.A., Friesch, D., Yu, Y., Main, D., Rambo, T., Simmons, J., Henry
 D., Wood, T.C., Leslie, A. and Wilkins, T.A.
 TITLE An integrated analysis of the genetics, development, and evolution
 JOURNAL of the cotton fiber
 COMMENT Unpublished (2000)
 CONTACT: Ming RA
 CLEMSON UNIVERSITY Genomics Institute
 100 Jordan Hall, Clemson, SC 29634, USA
 Tel: 864 656 7283
 Fax: 864 656 4293
 Email: twing@clemson.edu
 Seq primer: TAATTCGACTCAGCTATGAGG
 High quality sequence stop: 547.
 Location/Qualifiers
 1..552
 source
 /organism="Gossypium arboreum"
 /strain="AKA"
 /cultivar="8400"
 /db_xref="taxon:29729"
 /clone="GA_Eb0038F20f"
 /clone_lib="Gossypium arboreum 7-10 dpa fiber library"
 /rname_type="Fibers isolated from bolls harvested 7-10
 dpa"
 /lab_host="E. coli"

BASE COUNT 152 a 113 c 119 g 168 t
 ORIGIN
 Query Match 2.2%; Score 41.4; DB 12; Length 552;
 Best Local Similarity 49.3%; Pred. No. 9.1;
 Matches 108; Conservative 0; Mismatches 111; Indels 0; Gaps 0;
 QY 383 AATCTTTGCTTATCTGGCAGCTTACAGTATTTGATGCGCACTGCTTGACGA 442
 Db 62 AATATATGTTATGTTAGGTGAAGTGAAGCAAGCGGTTGTCAAAATGCTGATTTGGA 121
 QY 443 CTGCGTGGAGATGTTCTGTGTATGAGGATGCGGTAGATCAATGTTCTGATTCGA 502
 Db 122 CTGCAAGATTTATCAAGTCCCTTGAAGCCATTAATCTGAATATGTTGATTTACT 181
 QY 503 AGATCCGAGAGAAATTTTATGTCGCAAGCTTAAAAATCTGTAGAAAAGATAT 562
 Db 182 ATTTGGATCCGTCACCTGATTTGCTTGTGCAAGACATATACAGTCACTTGTAT 241
 QY 563 ACCAAGCTTTTGGAGCCATTTTGAATCTTAACTTGA 601
 Db 242 ATGTTGGCTGTTGATGATATTTTGTGCTGAGCTTTTGA 280

RESULT 38
 BU349244/c 676 bp mRNA linear EST 06-MAR-2002
 LOCUS
 DEFINITION BU349244 Dictyostelium discoideum cDNA library, AF Dictyostelium
 accession BU349244
 version BU349244.1 GI:19219751
 keywords
 source
 organism
 Dictyostelium discoideum.
 Eukaryota; Mycetozoa; Dictyostelidia; Dictyostelium.
 1 (bases 1 to 676)
 REFERENCE
 AUTHORS Urushihara, H., Tanaka, Y., Kohara, Y. and Shin-i, T.
 TITLE Full length cDNA of Dictyostelium discoideum at the aggregation
 JOURNAL stage
 COMMENT Unpublished (2002)
 CONTACT: Tadao Shin-i
 Center For Genetic Resource Information
 National Institute of Genetics
 1111 Yata, Mishima, Shizuoka 411-8540, Japan
 Tel: 81-559-81-6856
 Fax: 81-559-81-6855
 Email: tshin@genes.nig.ac.jp.
 Location/Qualifiers
 1..676
 source
 /organism="Dictyostelium discoideum"
 /strain="AX4"
 /db_xref="taxon:44689"
 /clone="dda35123"
 /clone_lib="Dictyostelium discoideum cDNA library, AF"
 /sex="mat A"
 /dev_stage="Aggregation stage"
 BASE COUNT 198 a 125 c 120 g 232 t 1 others
 ORIGIN
 Query Match 2.2%; Score 41.4; DB 13; Length 676;
 Best Local Similarity 50.0%; Pred. No. 9.8; Indels 0; Gaps 0;
 Matches 102; Conservative 0; Mismatches 102; Indels 0; Gaps 0;
 QY 836 GGAAGTCTTTTCTTTTATGCTGCTGCTGCGGTTGAGACCTGATTCGTTTG 895
 Db 210 GGTAAAGTACTGATGCTGATGCTTCAAAAGATTCGAAATCTCAAAACTTTAACT 151
 QY 896 GGAATGATTTTAAAGAGAGATGCTTTTACCTTTAATCAAAAGAGATGACATCAC 955
 Db 150 GCAAAATGCTTGGGTGTAAGGTACCGAAGTTTATGTTGTTAGGTGTTAGGTGTTAAAGTT 91
 QY 956 GATGTTGCTCAATGCTGGCAAGTTTGCAATTAACCTACAGGAAGCTGCTTTCTTCT 1015

ACCESSION BF295860
 VERSION BF295860.1 GI:13946305
 KEYWORDS EST.
 SOURCE Plasmodium berghei.
 ORGANISM Plasmodium berghei.
 REFERENCE 1 (bases 1 to 629)
 AUTHORS Carlton, J.M.-R. and Dame, J.B.
 TITLE The Plasmodium vivax and P. berghei gene sequence tag projects
 JOURNAL Parasitol. Today (Regul. Ed.) 16 (10), 409 (2000)
 COMMENT Contact: Dame JB
 Department of Pathobiology, College of Veterinary Medicine
 University of Florida
 2015 SW 23rd Avenue, Bldg 1017, Gainesville, FL 32611-0880, USA
 Tel: 352 392 4700
 Fax: 352 392 9704
 Email: damej@mail.vetmed.ufl.edu
 Seq primer: T3.

FEATURES
 source
 1. .629
 /organism="Plasmodium berghei"
 /strain="ANKA clone HP (gametocyte producer)"
 /db_xref="taxon:5821"
 /clone_lib="Pb cDNA #17, Tommaso Pace, Marta Ponzi, and Clara Frontali"
 /dev_stage="asynchronous blood stage"
 /lab_host="Wistar rats"
 /note="Vector: pBluescript II vector DNA, excised from Lambda ZAP II.; Site 1: EcoRI; Site 2: XhoI; Total RNA was extracted from asynchronous blood stage forms of the cloned ANKA isolate of P. berghei, grown in wistar rats to 30% parasitemia and 2-5% gametocytemia. Contaminating host cell cells had previously been removed and final host cell contamination estimated to be approximately 5%. PolyA+ RNA was extracted and reverse transcribed using an oligo dT-XhoI primer (Lambda ZAP II cDNA cloning kit, Stratagene). Second strand cDNA was made following the manufacturer's protocol. EcoRI adaptors were ligated to the cDNA, and fragments were ligated into EcoRI/XhoI digested vector."

BASE COUNT 251 a 47 c 76 g 252 t 3 others
 ORIGIN

Query Match 2.2%; Score 41.2; DB 12; Length 629;
 Best Local Similarity 49.3%; Pred. No. 11;
 Matches 103; Conservative 0; Mismatches 106; Indels 0; Gaps 0;

Qy 1606 AGTTTGTAAATGCTTTGTTATAGCGGCTCCTCTCTTAAATTTGCAATTTATAT 1665
 ||||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 616 AGTTTGTAAATGCTTTGTTATAGCGGCTCCTCTCTTAAATTTGCAATTTATAT 557
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Qy 1666 GACCATAGGATTTCTTAGGAACCTTTATCGTCTCTTATATTGCACCACTCTCTGTT 1725
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 556 AATATACAAATCATNTAAGCTTATTAATTTCTTATATTTCTTCAACCTTAATTTT 497
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Qy 1726 GTTATGTCGTAAGAAATTCGCTCAAAATAGTACCGTTAAATCTTAATCAACGTGT 1785
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 496 TTATACCATCACTACATATTTATTTTAAATAAAATGCATAATAACAATAAAATTG 437
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Qy 1786 AGCATATAAAATCTCTTTGGGACTTT 1814
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 436 CATAAAAAAACTAGTCCTCTTTTCTTTT 408
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 44
 CNS0111D
 LOCUS CNS0111D/c 750 bp DNA linear GSS 26-JUL-1999
 DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC
 BACN06D21 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.
 ACCESSION AL100303
 VERSION AL100303.1 GI:5611914
 KEYWORDS GSS.

SOURCE Drosophila melanogaster.
 ORGANISM Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 REFERENCE 1 (bases 1 to 750)
 AUTHORS Genoscope.
 TITLE Direct Submission
 JOURNAL Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
 COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaud at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBelOBAC11.

FEATURES
 Location/Qualifiers
 source
 1. .750
 /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"
 /clone="BACN06D21"
 /clone_lib="DrosBAC"
 /plasmid="pBelOBAC11"
 /notes="end : T7"

BASE COUNT 309 a 104 c 111 g 102 t 124 others
 ORIGIN

Query Match 2.2%; Score 41.2; DB 17; Length 750;
 Best Local Similarity 34.9%; Pred. No. 11;
 Matches 97; Conservative 44; Mismatches 137; Indels 0; Gaps 0;

Qy 621 TTCTTTCTCTAGATACAGGCGCTATTAAAGGTTGCTTGCATTCGATTTTAGGAA 680
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 386 TTTTITTTTWTCTTWTCTAAGTTKTYTAGTTTCTTATTTCTTAAKRWAKTTT 327
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Qy 681 TTTTCTCTCAATGTTTACGGCTCTTTTCATGACTAAATTTTCTTCATGCTGGATGA 740
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 326 TTTTCTCTAGYTTTWTGTTTCTTTTCTGTAATKTKTKTKTKTKTKTKTKTK 267
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Qy 741 ATAAGACCAACATACACAGTTGCATATGATGAATAGTTCGTGGGATAAAGCATGTT 800
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 266 KTAMTTTKKVTATATKATCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 207
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Qy 801 TCTTGAGAGGATCAAAAACCTTTGGGCTGTTTCTGGAAGTGTCTTTCTTTAGTTCG 860
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 206 WTTTWTAKTTTTCGAKKTTCTTTTWTYTKGTWKYKTTTGTGTTTWTWKTKT 147
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Qy 861 TTGCTCTCGGTTTGGAGCCTGGAATTCGGTTTGGGA 898
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 146 CTTTWTWKTKTATKGCCTGCAAGTCAGTTGTMGCA 109
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 45
 C92475/c
 LOCUS C92475 436 bp mRNA linear EST 12-JUL-1999
 DEFINITION C92475 Dictyostelium discoideum SS (H.urushihara) Dictyostelium
 discoideum CDNA clone SSE617, mRNA sequence.
 ACCESSION C92475
 VERSION C92475.1 GI:3074351
 KEYWORDS EST.
 SOURCE Dictyostelium discoideum.
 ORGANISM Dictyostelium discoideum
 Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
 REFERENCE 1 (bases 1 to 436)
 AUTHORS Morio, T., Urushihara, H., Saito, T., Ugawa, Y., Mizuno, H., Yoshida, M., Yoshino, R., Mieda, B.N., Pi, M., Sato, T., Takemoto, K., Yasukawa, H., Williams, J., Maeda, M., Takeuchi, I., Ochiai, H. and Tanaka, Y.
 TITLE The Dictyostelium developmental CDNA project: generation and analysis of expressed sequence tags from the first-finger stage of development

JOURNAL DNA Res. 5 (6), 335-340 (1998)
 MEDLINE 99156227
 COMMENT

Contact: Hideko Urushihara
 Institute of Biological Sciences
 University of Tsukuba
 1-1-1 Tennoudai, Tsukuba, Ibaraki 305-8572, Japan
 Tel: 81-298-53-4664
 Fax: 81-298-53-6614
 Email: hideko@biol.tsukuba.ac.jp
 PROJECT = 'Dictyostelium discoideum cDNA project in Japan'
 POLYA=No.

FEATURES
 source location/Qualifiers
 1. 436
 /organism="Dictyostelium discoideum"
 /strain="AX4"
 /db_xref="taxon:44689"
 /clone="SSE617"
 /dev_stage="slug"
 /dev_stage="slug"

BASE COUNT 228 a 36 c 54 g 118 t
 ORIGIN

Query Match 2.2% Score 41; DB 14; Length 436;
 Best Local Similarity 47.8%; Pred. No. 11; Mismatches 130; Indels 0; Gaps 0;
 Matches 119; Conservative 0; Mismatches 130; Indels 0; Gaps 0;
 QY 1608 TTTTGTATGCTTTTGTATATAGCGGCTCTGCTTAAATTTGCATTATATGA 1667
 Db 418 TTTTATATCTTTTCAATTTACCTGTTCTTTTCTTTTCTTTTCTTTTCTTTT 359
 QY 1668 CCATAGGATCTCTAGAACTTATGCTCTTTATATGACACCTCTGTTGT 1727
 Db 358 TCCGATTTGTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 299
 QY 1728 TTATGTCCTGAAGAAATCGCTCAAAATAAGTACGTTAACTTAATCTAACGTGAG 1787
 Db 298 TTTTATTAATTAATTAATTAACGAGAAAGAAATGAATTCATTTCAATTTT 239
 QY 1788 CAATATATAAATCTCTTGGACTTAGTCCAAAGGCCCTGTGATTAATTAATG 1847
 Db 238 GGTAAAAACGATTAATTAATTTTGTCAATGACAAATTTCTTTAAATTTTA 179
 QY 1848 ACAATTC 1856
 Db 178 TCCAAATTA 170

Search completed: December 17, 2002, 07:57:00
 Job time : 2017 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 17, 2002, 08:13:26 ; Search time 284 Seconds
(without alignments)
14780.721 Million cell updates/sec

Title: US-09-868-987-1
Perfect score: 1864
Sequence: 1 atggactccgcatattgtc.....atgacaaattcagataatgc 1864

Scoring table: OLIGO NUC
Gapop 60.0 , Gapext 60.0

Searched: 2185239 seqs, 1125999159 residues

Word size : 0

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : N_Geneseq 101002.*
1: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SID22/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SID22/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SID22/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SID22/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1864	100.0	1864	21	AAA50030
2	22	1.2	3256	20	AAV99971
3	21	1.1	570	24	ABQ33748
4	21	1.1	570	24	ABQ33749
5	21	1.1	573	24	ABQ32730
6	21	1.1	573	24	ABQ32731
7	20	1.1	20	20	AAV97008
8	20	1.1	20	20	AAV95807
9	20	1.1	20	20	AAV95808

					DNA encoding Chlam
					Nucleotide sequenc
					Oligonucleotide fo
					Oligonucleotide fo
					Oligonucleotide fo
					Oligonucleotide fo
					PCR primer used to
					PCR primer used to

C	10	20	1.1	20	20	AAV93334	PCR primer used to
	11	19	1.0	236	21	AAC05080	Human secreted pro
	12	19	1.0	420	21	AAC03937	Human secreted pro
C	13	19	1.0	439	22	AAK03932	Human brain expres
C	14	19	1.0	439	22	AAI13982	Probe #1915 for ge
C	15	19	1.0	439	22	AAI03846	Probe #1837 used t
C	16	19	1.0	445	22	AAK16882	Human brain expres
C	17	19	1.0	445	22	AAI233198	Probe #13131 for g
C	18	19	1.0	445	22	AAI08848	Probe #8839 used t
C	19	19	1.0	469	22	AAK50436	Human bone marrow
C	20	19	1.0	561	22	AAK37567	Human bone marrow
	21	19	1.0	812	22	AAH08684	Human cDNA clone (
	22	19	1.0	827	22	AAH05672	Human cDNA clone (
	23	19	1.0	1150	24	ABK35236	Human cDNA encodin
	24	19	1.0	1187	21	AAC81029	Human secreted pro
	25	19	1.0	1240	22	AAI58527	Human polynucleoti
	26	19	1.0	1311	22	AAH33729	Human colon cancer
C	27	19	1.0	1684	22	AAI60313	Human polynucleoti
	28	19	1.0	1684	22	AAH68588	Human protein HP10
	29	19	1.0	1685	22	AAH16819	Human cDNA sequenc
	30	19	1.0	1690	22	AAH16007	Human cDNA sequenc
	31	19	1.0	1749	24	AAD24382	Human RNA metaboli
	32	19	1.0	2364	20	AAV99969	Nucleotide sequenc
C	33	19	1.0	2364	24	AAI39688	CDNA encoding the
C	34	19	1.0	2972	19	AAV29125	Homo sapiens Smoot
C	35	19	1.0	2972	22	AAC65391	Human Smoothened c
	36	19	1.0	19082	24	ABL32627	Human immune syste
	37	19	1.0	22660	23	ABL09700	Drosophila melanog
C	38	19	1.0	31352	24	AAI39687	Genomic DNA encodi
	39	18	1.0	235	24	ABN26694	Human ORFX polynuc
C	40	18	1.0	357	22	AAI91598	Human polynucleoti
	41	18	1.0	378	24	ABN90719	Staphylococcus epi
C	42	18	1.0	459	24	ABK44466	CDNA encoding colo
C	43	18	1.0	478	24	ABL93238	Arabidopsis thalia
	44	18	1.0	576	23	ABV19019	Human prostate exp
	45	18	1.0	613	24	ABG66275	Arabidopsis thalia
	46	18	1.0	700	22	AAH92994	Human inflammatory
	47	18	1.0	700	22	AAH92995	Human inflammatory
	48	18	1.0	713	22	AAH90788	Human digestive sy
	49	18	1.0	713	23	ABK42473	Genomic sequence #
	50	18	1.0	790	21	AAI2029	Aspergillus oryzae
	51	18	1.0	873	22	AAH03313	Human cDNA clone (
	52	18	1.0	922	22	AAK90789	Human digestive sy
	53	18	1.0	922	23	ABK42474	Genomic sequence #
C	54	18	1.0	922	24	ABL34272	Human immune syste
C	55	18	1.0	922	24	ABL34632	Human metastasis a
	56	18	1.0	1233	23	AAH89429	DNA encoding novel
	57	18	1.0	1233	23	AAH93360	DNA encoding novel
C	58	18	1.0	1676	22	AAI59174	Human polynucleoti
	59	18	1.0	1713	22	AAH16971	Human cDNA sequenc
	60	18	1.0	1714	22	AAI35706	Human musculoskele
	61	18	1.0	1714	22	AAI35707	Human musculoskele
	62	18	1.0	2167	22	AAH44797	Cysteine protease
	63	18	1.0	2259	17	AAI30686	Kaposi's sarcoma a
	64	18	1.0	2259	17	AAI16816	Kaposi's sarcoma a
C	65	18	1.0	2330	19	AAI296367	S. pneumoniae deri
	66	18	1.0	2347	18	AAV74464	Staphylococcus aur
	67	18	1.0	2487	22	AAH79943	Nucleotide sequenc
	68	18	1.0	2487	22	AAH79945	Nucleotide sequenc
	69	18	1.0	2583	21	AAH98005	Murine T gene DNA
	70	18	1.0	2631	21	AAH50996	Arabidopsis thalia
	71	18	1.0	2658	21	AAH50885	S. epidermidis gen
	72	18	1.0	2902	22	AAH54806	S. epidermidis gen
	73	18	1.0	2916	23	AAH92744	DNA encoding novel
	74	18	1.0	3222	23	AAH88366	DNA encoding novel
C	75	18	1.0	3222	23	AAH92786	DNA encoding novel
C	76	18	1.0	3350	22	AAH54654	S. epidermidis gen
	77	18	1.0	3598	23	AAH86420	DNA encoding novel
	78	18	1.0	3692	18	AAH72107	Rat semaphorin 2 g
	79	18	1.0	3715	21	AAH258781	Human HIP1 exon 29
	80	18	1.0	3810	23	ABL28301	Drosophila melanog
	81	18	1.0	3885	23	AAH77365	DNA encoding novel
	82	18	1.0	4796	18	AAH67187	Huntingtin interac

83	18	1.0	4796	21	AAZ58746	Human huntingtin-i
84	18	1.0	5710	24	ABJ33622	Human immune syste
85	18	1.0	6259	22	ABA08141	Human ovarian and
86	18	1.0	6259	22	AAJ07001	Human reproductive
87	18	1.0	6260	22	ABA08142	Human ovarian and
88	18	1.0	6260	22	AAJ07002	Human reproductive
89	18	1.0	6564	24	ABJ32123	Human immune syste
90	18	1.0	6874	23	ABJ09960	Drosophila melanog
91	18	1.0	6882	23	ABJ20434	Drosophila melanog
92	18	1.0	7237	23	ABJ28300	Drosophila melanog
93	18	1.0	7754	23	ABJ15284	Drosophila melanog
94	18	1.0	8144	20	AAK32482	Porcine adipocyte
95	18	1.0	13279	23	ABJ20172	Drosophila melanog
96	18	1.0	15656	23	ABJ09968	Methanococcus jann
97	18	1.0	16550	19	AAV21211	Kaposi's sarcoma a
98	18	1.0	20710	17	AAJ30681	Kaposi's sarcoma a
99	18	1.0	20710	17	AAJ30680	Streptococcus pneu
100	18	1.0	28473	19	AAV52216	KSJV LTR DNA (nucl
101	18	1.0	35100	20	AAV73803	Human kinase gene.
102	18	1.0	36159	20	ABN85329	V. marinus PKS-lik
103	18	1.0	40138	20	AAV81946	Human T gene DNA.
104	18	1.0	40138	21	AAV71520	Human T gene DNA.
105	18	1.0	81369	21	AAV7997	KSJV long unique c
106	18	1.0	137507	19	AAV19941	Soybean 240017 reg
107	18	1.0	335913	22	AAJ61371	Soybean 240017 reg
108	18	1.0	335913	22	AAJ61372	Pyrococcus abyssi
109	18	1.0	349980	22	AAJ66431	Lacteria innocua c
110	18	1.0	1163020	24	ABO67197	Lacteria innocua D
111	18	1.0	3011208	24	ABO66245	Human uteroglobin
112	17	0.9	33	24	ABJ54065	Human spliced tran
113	17	0.9	60	24	ABN33194	Human secreted pro
114	17	0.9	105	21	AAJ26327	Gene expression pr
115	17	0.9	150	22	AAJ04576	Toxicologically re
116	17	0.9	150	24	ABJ96661	Human secreted pro
117	17	0.9	160	21	AAJ05249	Human secreted pro
118	17	0.9	162	20	AAJ40273	Human foetal liver
119	17	0.9	178	22	ABA76485	Probe #19465 for g
120	17	0.9	178	22	ABA40999	Human bone marrow
121	17	0.9	178	22	AAK25138	Human brain expres
122	17	0.9	178	22	AAK51138	Human bone marrow
123	17	0.9	178	22	AAJ57171	Probe #25857 used
124	17	0.9	178	22	ABJ24659	Human genome-deriv
125	17	0.9	256	24	ABN17025	Human ORFX polynuc
126	17	0.9	281	22	AAJ400113	DNA encoding human
127	17	0.9	281	22	AAJ00689	Human reproductive
128	17	0.9	295	20	AAJ87548	Human single nucle
129	17	0.9	296	24	ABJ37798	Human colon tumour
130	17	0.9	301	24	ABN18829	Human ORFX polynuc
131	17	0.9	313	16	AAJ19150	Human gene signatu
132	17	0.9	316	23	AAJ78195	DNA encoding novel
133	17	0.9	324	24	ABJ64574	Stomach cancer rel
134	17	0.9	324	24	ABJ64574	Stomach cancer rel
135	17	0.9	373	23	ABJ08285	Human prostate exp
136	17	0.9	374	21	AAJ08781	Human secreted pro
137	17	0.9	384	22	AAJ38911	Novel human diagno
138	17	0.9	403	24	ABJ83175	Human ovarian cance
139	17	0.9	410	22	AAJ67353	Novel human polynu
140	17	0.9	412	22	AAJ84772	Novel human polynu
141	17	0.9	413	22	AAJ67213	Novel human polynu
142	17	0.9	421	22	AAJ90149	Human polynucleoti
143	17	0.9	425	23	ABJ38193	Human prostate exp
144	17	0.9	434	23	ABJ09466	Human prostate exp
145	17	0.9	450	23	ABJ30642	Human prostate exp
146	17	0.9	450	23	ABJ30642	Human prostate exp
147	17	0.9	459	23	ABJ20231	Human prostate exp
148	17	0.9	459	23	ABJ20231	Human prostate exp
149	17	0.9	473	22	AAJ16715	Human breast cance
150	17	0.9	478	24	ABJ82612	Human ovarian cance
151	17	0.9	480	24	ABJ7386	Streptococcus poly
152	17	0.9	485	23	AAJ86415	DNA encoding novel
153	17	0.9	490	21	AAJ79485	Eucalyptus grandis
154	17	0.9	492	21	AAJ48108	Zea mays DNA fragm
155	17	0.9	492	21	AAJ3501	Zea mays DNA fragm
156	17	0.9	494	21	AAJ44313	Arabidopsis thalia
157	17	0.9	494	22	AAJ57170	Human foetal liver
158	17	0.9	494	22	AAJ05198	Human brain expres
159	17	0.9	494	22	AAJ30786	Human bone marrow
160	17	0.9	494	22	AAJ36699	Probe #5385 used t
161	17	0.9	494	24	ABJ05531	Human genome-deriv
162	17	0.9	496	22	AAJ88741	Invertebrate cDNA se
163	17	0.9	500	24	ABN63137	Human cancer relat
164	17	0.9	528	23	AAJ86414	DNA encoding novel
165	17	0.9	538	22	AAJ11097	Human cDNA clone (
166	17	0.9	539	22	AAJ61076	Human foetal liver
167	17	0.9	539	22	AAJ28978	Probe #7444 for ge
168	17	0.9	539	22	AAJ09369	Human brain expres
169	17	0.9	539	22	AAJ35258	Human bone marrow
170	17	0.9	539	22	AAJ40973	Probe #9659 used t
171	17	0.9	539	22	ABJ09672	Human genome-deriv
172	17	0.9	543	21	AAJ12591	DNA encoding the o
173	17	0.9	549	21	AAJ39761	Cat flea hindgut a
174	17	0.9	549	22	AAJ88740	Invertebrate cDNA se
175	17	0.9	556	24	ABN18003	Human ORFX polynuc
176	17	0.9	558	20	AAJ21087	Polynucleotide seq
177	17	0.9	559	23	AAJ86416	DNA encoding novel
178	17	0.9	565	24	ABJ50492	Human nGPCR-Seg107
179	17	0.9	595	18	AAJ75439	Staphylococcus aur
180	17	0.9	599	22	AAJ64479	Novel human polynu
181	17	0.9	599	24	ABJ69949	Pancreas cancer re
182	17	0.9	605	23	ABJ50707	Human prostate exp
183	17	0.9	607	23	ABJ55343	Human prostate exp
184	17	0.9	613	24	ABJ35658	Oligonucleotide fo
185	17	0.9	613	24	ABJ35659	Oligonucleotide fo
186	17	0.9	615	24	ABJ99824	Human secretory po
187	17	0.9	622	24	AAJ45843	Human secreted pro
188	17	0.9	643	21	AAJ11257	Aspergillus niger
189	17	0.9	676	21	AAJ14103	Aspergillus oryzae
190	17	0.9	700	22	AAJ92072	Human inflammatory
191	17	0.9	700	22	AAJ92324	Human inflammatory
192	17	0.9	700	22	AAJ92325	Human inflammatory
193	17	0.9	700	22	AAJ92429	Human inflammatory
194	17	0.9	700	22	AAJ92430	Human inflammatory
195	17	0.9	721	20	AAJ20471	Human secreted pro
196	17	0.9	743	24	AAJ46509	M catarrhalis MC41
197	17	0.9	747	18	AAJ74961	Staphylococcus aur
198	17	0.9	800	19	AAJ27566	Human glutathione
199	17	0.9	820	24	ABJ34510	Oligonucleotide fo
200	17	0.9	820	24	ABJ34511	Oligonucleotide fo
201	17	0.9	829	22	AAJ85953	Human polynucleoti
202	17	0.9	917	24	ABJ069105	Lacteria monocytog
203	17	0.9	945	22	AAJ31818	Human olfactory re
204	17	0.9	948	21	AAJ293817	Open reading frame
205	17	0.9	948	21	AAJ293819	Open reading frame
206	17	0.9	948	21	AAJ293819	Human DNA for olfa
207	17	0.9	952	24	ABJ08053	Human GPCRX cDNA #
208	17	0.9	971	21	AAJ56349	Pinus radiata tran
209	17	0.9	979	22	AAJ88788	Invertebrate cDNA seq
210	17	0.9	1137	20	AAJ25024	Murine D6 encoding
211	17	0.9	1208	21	AAJ81594	N. meningitidis pa
212	17	0.9	1219	23	ABJ03281	Drosophila melanog
213	17	0.9	1263	24	ABN68910	Streptococcus poly
214	17	0.9	1263	24	ABN70499	Streptococcus poly
215	17	0.9	1273	21	AAJ14181	Aspergillus oryzae
216	17	0.9	1300	21	AAJ47348	Arabidopsis thalia
217	17	0.9	1301	24	ABJ55524	Human ATP-depend
218	17	0.9	1304	21	AAJ33192	Arabidopsis thalia
219	17	0.9	1323	21	AAJ095484	Streptococcus poly
220	17	0.9	1332	22	AAJ34187	Human secreted pro
221	17	0.9	1353	22	AAJ34187	Human colon cancer
222	17	0.9	1355	20	AAJ29276	Human prostate tum
223	17	0.9	1366	22	AAJ14142	Human cDNA sequenc
224	17	0.9	1468	22	AAJ69132	Human immune/haema
225	17	0.9	1484	14	AAJ048745	Al4 self incompati
226	17	0.9	1490	21	AAJ291252	Plasmid LPD-LBI-II
227	17	0.9	1490	22	ABJ14571	Human nervous syst
228	17	0.9	1490	22	ABJ19232	Human nervous syst

5. epidermidis gen	PR	23-DEC-1998;	98US-0113281.
DNA encoding novel	PR	23-DEC-1998;	98US-0113282.
Human prostate can	PR	23-DEC-1998;	98US-0113283.
Human DNA sequence	PR	23-DEC-1998;	98US-0113284.
Human tumour-associ	PR	23-DEC-1998;	98US-0113285.
Human G protein as	PR	23-DEC-1998;	98US-0113285.
Human colon speci	PR	28-DEC-1998;	98US-0114050.
Human prostate exp	PR	28-DEC-1998;	98US-0114056.
Drosophila melanog	PR	28-DEC-1998;	98US-0114057.
Human secreted pro	PR	28-DEC-1998;	98US-0114058.
Drosophila melanog	PR	28-DEC-1998;	98US-0114059.
Partial sequence o	PR	28-DEC-1998;	98US-0114061.
Human breast cance	XX		
Human breast cance	PA	(CONN-) CONNAUGHT LAB LTD.	
Drosophila melanog	XX		
Drosophila melanog	XX		
Drosophila melanog	XX		
Drosophila melanog	XX		
Drosophila melanog	XX		
Human cancer agent	XX		
Drosophila melanog	XX		
DNA encoding novel	PT		
Drosophila melanog	PT		
Human cervical can	XX		
Human receptor-tyr	PS		
Drosophila melanog	XX		
Drosophila melanog	CC		
Drosophila melanog	CC		
Drosophila melanog	CC		
DNA encoding novel	CC		
Drosophila melanog	CC		
Drosophila melanog	CC		
Arabidopsis thaila	CC		
Human nervous syst	CC		
Human nervous syst	CC		
Human immune/haema	CC		
Human nervous syst	CC		
Human nervous syst	CC		
Human immune/haema	CC		
Human immune/haema	CC		
T. thermophila tri	XX		
Genomic sequence #	XX		
Genomic sequence #	XX		
Restriction resistance	XX		

Query Match	100.0%;	Score 1864;	DB 21;	Length 1664;
Best Local Similarity	100.0%;	Pred. No. 0;		
Sequence 1864 BP; 498 A; 346 C; 401 G; 619 T; 0 other;				

Claim 2(a); Page 55-58; 215pp; English.

The present sequence is that of Chlamydia pneumoniae genomic DNA including an open reading frame that codes for CPN100586 RY 54 (see AA95543), a putative 93 kDa outer membrane protein. It is an example of C. pneumoniae polynucleotide molecules of the invention (see AA50030-42) that encode antigenic polypeptides (see AA95543-55) useful in the diagnosis, treatment and prevention of Chlamydia infection. The polynucleotides can be utilised: in the recombinant production of Chlamydia antigens using transformed intracellular host cells; in vaccines or live vaccine vectors; in naked form or formulated with a delivery vehicle for therapy and prophylaxis of Chlamydia infection; and as probes or primers for diagnosis of Chlamydia infection.

1 ATGGACTTCCGCATATTGTCAGGAGGGGATCAGCGGCACTGCTAATGGACATATTCTGC 60

PR 23-DEC-1998; 98US-0113280.

XX	PR	23-DEC-1998;	98US-0113261.
XX	PR	23-DEC-1998;	98US-0113282.
XX	PR	23-DEC-1998;	98US-0113283.
XX	PR	23-DEC-1998;	98US-0113284.
XX	PR	23-DEC-1998;	98US-0113285.
XX	PR	23-DEC-1998;	98US-0113385.
XX	PR	28-DEC-1998;	98US-0114050.
XX	PR	28-DEC-1998;	98US-0114056.
XX	PR	28-DEC-1998;	98US-0114057.
XX	PR	28-DEC-1998;	98US-0114058.
XX	PR	28-DEC-1998;	98US-0114059.
XX	PR	28-DEC-1998;	98US-0114061..
PA	(CONN-) CONNAUGHT LAB LTD.		
PI	Murdin AD, Oomen RP, Wang J;		
XX	WPI; 2000-452369/39.		
DR	P-SDB; AAY95543.		
XX			
PT	Novel Chlamydia polynucleotides and polypeptides useful for diagnosis,		
PT	prevention and treatment of Chlamydia infection in mammals -		
PS	Claim 2(a); Page 55-58; 215pp; English.		
XX			
CC	The present sequence is that of Chlamydia pneumoniae genomic DNA		
CC	including an open reading frame that codes for CPN10066 RY 54 (see		
CC	AAY95543), a putative 93 kDa outer membrane protein. It is an example		
CC	of C. pneumoniae polynucleotide molecules of the invention (see		
CC	AA50030-42) that encode antigenic polypeptides (see AAY9543-55) useful		
CC	in the diagnosis, treatment and prevention of Chlamydia infection.		
CC	The polynucleotides can be utilised: in the recombinant production		
CC	of Chlamydia antigens using transformed unicellular host cells; in		
CC	vaccines or live vaccine vectors; in naked form or formulated with		
CC	a delivery vehicle for therapy and prophylaxis of Chlamydia		
CC	infection; and as probes or primers for diagnosis of Chlamydial		
CC	infection.		
SQ	Sequence 1864 BP; 498 A; 346 C; 401 G; 619 T; 0 other;		
Query Match	100.0%; Score 1864; DB 21; Length 1864;		
Best Local Similarity	100.0%; Pred. No. 0;		
Matches 1864; Conservative	0; Mismatches 0; Indels 0; Gaps 0;		
OY	1 ATGACATTCGGCATATTGTCTCAGAGGGGATCAGCGGCATCTGTAATGACAATAATTTCTGC	60	
Db	1 ATGACATTCGGCATATTGTCTCAGAGGGGATCAGCGGCATCTGTAATGACAATAATTTCTGC	60	
OY	61 AAACCGTGANNGGGTANNGGCTGATGATGAAGGTTAATGTCAGCACGCCCATTTT	120	
Db	61 AAACCGTGANNGGGTANNGGCTGATGATGAAGGTTAATGTCAGCACGCCCATTTT	120	
OY	121 AACGTCOCATTTGAAAAATCATGCGCAGTGTCTCAGGGAAATTTACCACCGTGAAGTAG	180	
Db	121 AACGTCOCATTTGAAAAATCATGCGCAGTGTCTCAGGGAAATTTACCACCGTGAAGTAG	180	
OY	181 CAAACTGGCTCAGATTAAAATCTGAGCGCATGCTTTTGTCGCGAGGTTCTCACTGA	240	
Db	181 CAAACTGGCTCAGATTAAAATCTGAGCGCATGCTTTTGTCGCGAGGTTCTCACTGA	240	
OY	241 AGAAGCATCTCTTCTGTGATCTTGGGAAAAAACATGTCACAAAGCATTAATCTCAGCATG	300	
Db	241 AGAAGCATCTCTTCTGTGATCTTGGGAAAAAACATGTCACAAAGCATTAATCTCAGCATG	300	
OY	301 CTGAGGCTTGSCANTGCTTATGTTTTGATGAGGATATTAATGATTTGGAGGCGCAT	360	
Db	301 CTGAGGCTTGSCANTGCTTATGTTTTGATGAGGATATTAATGATTTGGAGGCGCAT	360	
OY	361 CGCTTCGGAGAGCTTTCTTCTGAACTTTTGGCTTAATCTAGGCGAGCTTCACGATATTTGA	420	
Db	361 CGCTTCGGAGAGCTTTCTTCTGAACTTTTGGCTTAATCTAGGCGAGCTTCACGATATTTGA	420	
OY	421 TGCGCACTACCTTTGTCAGACTGCTGGGATGTTTCTTGCTATGGGATGCGCGTGA	480	

PR 21-JUL-1997; 97US-0897798.
 PR 01-JUL-1997; 97US-0051407.
 XX
 XX (INGH/) INGHAM P W.
 PA (ONTO-) ONTOGENY INC.
 PA (VHEU/) VAN DEN HEUVEL M.
 XX
 XX WPI; 1999-105996/09.
 DR P-PSDB; AAW81066.
 XX
 XX
 PT New isolated smoothened genes - used to develop products for
 PT treating e.g. cancers, neurodegenerative disorders, nervous system
 PT injury, immunological diseases or infections
 XX
 XX
 PS Claim 12; Pages 93-94; 107pp; English.

CC This is the nucleotide sequence of the novel chicken smoothened
 CC gene used in the method of the invention to produce smoothened
 CC polypeptides (SPs). The SPs can modulate at least one of
 CC proliferation, differentiation or survival of a cell which
 CC expresses the SP. The products can be used to modulate
 CC spermatogenesis, osteogenesis, chondrogenesis or neuronal cell
 CC differentiation or to enhance survival of neuronal cells, e.g. to
 CC prevent apoptosis. They can be used for treating e.g. cancer,
 CC degenerative disorders marked by loss of particular cell-types,
 CC apoptosis, neoplastic and/or hyperplastic disorders. In particular
 CC they can be used to treat neurological conditions derived from
 CC acute, subacute, or chronic injury to the nervous system, aging of
 CC the nervous system chronic neurodegenerative diseases of the
 CC nervous system, chronic immunological diseases of the nervous
 CC system or affecting the nervous system. They can also treat be used
 CC to CNS trauma infection, infection (such as viral infection with
 CC varicella-zoster), metabolic disease, nutritional deficiency,
 CC toxic agents (such as cisplatin treatment), chronic pain syndromes
 CC or in nerve prostheses for the repair of central and peripheral
 CC nerve damage, malignant gliomas, medulloblastomas, neuroectodermal
 CC tumours, ependymomas and for liver repair and regeneration of lung
 CC tissue in the treatment of emphysema, or for enhancing bond
 CC formation, tissue formation or fertility. The products can also be
 CC used for detection, diagnosis and drug screening.

XX Sequence 3256 BP; 594 A; 1015 C; 953 G; 669 T; 25 other;

Query Match 1.2%; Score 22; DB 20; Length 3256;
 Best Local Similarity 100.0%; Pred. No. 1.2;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1277 GCTTTGGCAATCATCTTGTCT 1298
 |||||
 DB 1697 GCTTTGGCAATCATCTTGTCT 1676

RESULT 3
 ABQ33748/C
 ID ABQ33748 standard; DNA; 570 BP.
 XX
 AC ABQ33748;
 XX

DT 12-JUL-2002 (first entry)
 XX
 DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 20339.
 XX
 XX Human: cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
 KM drug; side effect; cancer; central nervous system; cardiovascular;
 KM gastrointestinal; respiratory system; single nucleotide polymorphism;
 KM SNP; cell differentiation; ds.
 XX
 XX Homo sapiens.
 OS
 XX
 PN WO200218632-A2.
 XX
 PD 07-MAR-2002.

PF 01-SEP-2001; 2001WO-EP10074.
 XX
 XX 01-SEP-2000; 2000DE-1043826.
 PR 05-SEP-2000; 2000DE-1044543.
 XX
 XX (EPIC-) EPIGENOMICS AG.
 PA
 PI Olek A, Piepenbrock C, Berlin K, Guetig D;
 XX
 XX WPI; 2002-371829/40.
 DR
 XX
 XX
 PT Determining the degree of cytosine methylation in genomic DNA, useful
 PT for diagnosis and prognosis, comprises selective hybridization of
 PT amplicons from chemically treated DNA -
 XX
 XX
 PS Claim 12; 56pp + Sequence Listing; 56pp; German.

CC This invention describes a novel method for determining the degree of
 CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
 CC genomic sample of DNA. The sample is treated chemically to convert in a
 CC cytosine (C) but not methylated C, to uracil, then part of the genomic
 CC DNA that contains the target C is amplified to form a labeled amplicon.
 CC The amplicon is hybridised to two classes, each with at least one
 CC member of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
 CC and the degree of hybridisation to both classes is determined from the
 CC label on the amplicon. From the ratio of labels hybridised to the two
 CC classes of oligomers, the degree of methylation is calculated. The method
 CC is used: (i) for diagnosis and/or prognosis of side effects of
 CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
 CC of the central nervous, cardiovascular, gastrointestinal and respiratory
 CC systems etc., particularly by detecting mutations or single nucleotide
 CC polymorphisms (SNPs); and (ii) for differentiation of cell or tissue
 CC types and for investigating cell differentiation. The method allows the
 CC methylation status of many C residues to be determined simultaneously.
 CC ABQ33410-ABQ54121 represent genomic DNA sequences used to illustrate the
 CC method for determining the degree of cytosine methylation described in
 CC the disclosure of the invention.

XX Sequence 570 BP; 93 A; 53 C; 218 G; 206 T; 0 other;

Query Match 1.1%; Score 21; DB 24; Length 570;
 Best Local Similarity 100.0%; Pred. No. 3.7;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1185 GAAACTCTAAGCAACGCA 1205
 |||||
 DB 287 GAAACTCTAAGCAACGCA 267

RESULT 4
 ABQ33749
 ID ABQ33749 standard; DNA; 570 BP.
 XX
 AC ABQ33749;
 XX

DT 12-JUL-2002 (first entry)
 XX
 DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 20340.
 XX
 XX Human: cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
 KM drug; side effect; cancer; central nervous system; cardiovascular;
 KM gastrointestinal; respiratory system; single nucleotide polymorphism;
 KM SNP; cell differentiation; ds.
 XX
 XX Homo sapiens.
 OS
 XX
 PN WO200218632-A2.
 XX
 PD 07-MAR-2002.
 XX
 PF 01-SEP-2001; 2001WO-EP10074.
 XX
 PR 01-SEP-2000; 2000DE-1043826.

PR 05-SEP-2000; 2000DE-1044543.
 XX (EPIG-) EPIGENOMICS AG.
 PI Olek A, Piepenbrock C, Berlin K, Guetig D;
 XX WPI; 2002-371829/40.
 XX Determining the degree of cytosine methylation in genomic DNA, useful
 PT for diagnosis and prognosis, comprises selective hybridization of
 PT amplicons from chemically treated DNA -
 XX Claim 12; 56pp + Sequence Listing; 56pp; German.
 XX This invention describes a novel method for determining the degree of
 CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
 CC genomic sample of DNA. The sample is treated chemically to convert
 CC cytosine (C) but not methylated C, to uracil, then part of the genomic
 CC DNA that contains the target C is amplified to form a labeled amplicon.
 CC The amplicon is hybridised to two classes, each with at least one
 CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
 CC and the degree of hybridisation to both classes is determined from the
 CC label on the amplicon. From the ratio of labels hybridised to the two
 CC classes of oligomers, the degree of methylation is calculated. The method
 CC is used: (i) for diagnosis and/or prognosis of side effects of
 CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
 CC of the central nervous, cardiovascular, gastrointestinal and respiratory
 CC systems etc., particularly by detecting mutations or single nucleotide
 CC polymorphisms (SNPs); and (ii) for differentiation of cell or tissue
 CC types and for investigating cell differentiation. The method allows the
 CC methylation status of many C residues to be determined simultaneously.
 CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
 CC method for determining the degree of cytosine methylation described in
 CC the disclosure of the invention.
 XX Sequence 570 BP; 206 A; 218 C; 53 G; 93 T; 0 other;
 SQ
 Query Match 1.1%; Score 21; DB 24; Length 570;
 Best Local Similarity 100.0%; Pred. No. 3.7;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1185 GAAACTCTAAACGAAACGCA 1205
 DB 284 GAAACTCTAAACGAAACGCA 304
 RESULT 5
 ABQ32730/C
 ID ABQ32730 standard; DNA; 573 BP.
 XX AC ABQ32730;
 XX 12-JUL-2002 (first entry)
 XX Oligonucleotide for detecting cytosine methylation SEQ ID NO 19321.
 XX Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
 XX drug; side effect; cancer; central nervous system; cardiovascular;
 XX gastrointestinal; respiratory system; single nucleotide polymorphism;
 XX SNP; cell differentiation; ds.
 XX Homo sapiens.
 XX WO200218632-A2.
 XX 07-MAR-2002.
 XX 01-SEP-2001; 2001WO-EP10074.
 XX 01-SEP-2000; 2000DE-1043826.
 XX 05-SEP-2000; 2000DE-1044543.
 XX (EPIG-) EPIGENOMICS AG.

XX Olek A, Piepenbrock C, Berlin K, Guetig D;
 XX WPI; 2002-371829/40.
 XX Determining the degree of cytosine methylation in genomic DNA, useful
 PT for diagnosis and prognosis, comprises selective hybridization of
 PT amplicons from chemically treated DNA -
 XX Claim 12; 56pp + Sequence Listing; 56pp; German.
 XX This invention describes a novel method for determining the degree of
 CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
 CC genomic sample of DNA. The sample is treated chemically to convert
 CC cytosine (C) but not methylated C, to uracil, then part of the genomic
 CC DNA that contains the target C is amplified to form a labeled amplicon.
 CC The amplicon is hybridised to two classes, each with at least one
 CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
 CC and the degree of hybridisation to both classes is determined from the
 CC label on the amplicon. From the ratio of labels hybridised to the two
 CC classes of oligomers, the degree of methylation is calculated. The method
 CC is used: (i) for diagnosis and/or prognosis of side effects of
 CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
 CC of the central nervous, cardiovascular, gastrointestinal and respiratory
 CC systems etc., particularly by detecting mutations or single nucleotide
 CC polymorphisms (SNPs); and (ii) for differentiation of cell or tissue
 CC types and for investigating cell differentiation. The method allows the
 CC methylation status of many C residues to be determined simultaneously.
 CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
 CC method for determining the degree of cytosine methylation described in
 CC the disclosure of the invention.
 XX Sequence 573 BP; 82 A; 49 C; 229 G; 207 T; 6 other;
 SQ
 Query Match 1.1%; Score 21; DB 24; Length 573;
 Best Local Similarity 100.0%; Pred. No. 3.7;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1185 GAAACTCTAAACGAAACGCA 1205
 DB 200 GAAACTCTAAACGAAACGCA 180
 RESULT 6
 ABQ32731
 ID ABQ32731 standard; DNA; 573 BP.
 XX AC ABQ32731;
 XX 12-JUL-2002 (first entry)
 XX Oligonucleotide for detecting cytosine methylation SEQ ID NO 19322.
 XX Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
 XX drug; side effect; cancer; central nervous system; cardiovascular;
 XX gastrointestinal; respiratory system; single nucleotide polymorphism;
 XX SNP; cell differentiation; ds.
 XX Homo sapiens.
 XX WO200218632-A2.
 XX 07-MAR-2002.
 XX 01-SEP-2001; 2001WO-EP10074.
 XX 01-SEP-2000; 2000DE-1043826.
 XX 05-SEP-2000; 2000DE-1044543.
 XX (EPIG-) EPIGENOMICS AG.
 XX Olek A, Piepenbrock C, Berlin K, Guetig D;
 XX

DR WPI; 2002-371829/40.

XX Determining the degree of cytosine methylation in genomic DNA, useful

PT for diagnosis and prognosis, comprises selective hybridization of

PT amplicons from chemically treated DNA -

XX

PS Claim 12; 56pp + Sequence listing; 56pp; German.

XX This invention describes a novel method for determining the degree of

CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a

CC genomic sample of DNA. The sample is treated chemically to convert

CC cytosine (C) but not methylated C, to uracil, then part of the genomic

CC DNA that contains the target C is amplified to form a labeled amplicon.

CC The amplicon is hybridised to two classes, each with at least one

CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers

CC and the degree of hybridisation to both classes is determined from the

CC label on the amplicon. From the ratio of labels hybridised to the two

CC classes of oligomers, the degree of methylation is calculated. The method

CC is used: (i) for diagnosis and/or prognosis of side effects of

CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders

CC of the central nervous, cardiovascular, gastrointestinal and respiratory

CC systems etc., particularly by detecting mutations or single nucleotide

CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue

CC types and for investigating cell differentiation. The method allows the

CC methylation status of many C residues to be determined simultaneously.

CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the

CC method for determining the degree of cytosine methylation described in

CC the disclosure of the invention.

CC

XX

SQ Sequence 573 BP; 207 A; 229 C; 49 G; 82 T; 6 other;

Query Match 1.1%; Score 21; DB 24; Length 573;

Best Local Similarity 100.0%; Pred.No.3.7;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1185 GAAACTCTAAACGAAACGCAA 1205

DB 374 GAAACTCTAAACGAAACGCAA 394

RESULT 7

AAK97008/C

ID AAK97008 standard; DNA; 20 BP.

XX

AC AAK97008;

XX

DT 13-SEP-1999 (first entry)

XX

DE PCR primer used to amplify an ORF of Chlamydia pneumoniae.

XX

KW Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;

KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;

KW vaccine; neutralising epitope; PCR primer; ss.

XX

OS Synthetic.

OS Chlamydia pneumoniae.

XX

PN WO927105-A2.

XX

PD 03-JUN-1999.

XX

PF 20-NOV-1998; 98WO-IB01890.

XX

PR 04-NOV-1998; 98US-0107078.

PR 21-NOV-1997; 97FR-0014673.

XX

PA (GEST) GENSET.

XX

PI Griffais R;

XX

DR WPI; 1999-357842/30.

XX

PT Genome sequence of Chlamydia pneumoniae

XX

PS Page 1870; Disclosure; 1912pp; English.

XX

CC AAK91991-X97517 represent PCR primers used to amplify open reading

CC frames and other nucleic acid sequences from the genome of

CC Chlamydia pneumoniae (see AAK91990). C. pneumoniae causes respiratory

CC disease such as pneumonia and bronchitis and is thought to be a

CC contributing factor in heart disease, sarcoidosis, sinusitis, purulent

CC otitis media, erythema nodosum or pharyngitis. The polypeptides encoded

CC by the open reading frames of the C. pneumoniae genome (see AAY34584-

CC AAY35879) can be used in immunogenic compositions as vaccines. Vectors

CC containing C. pneumoniae nucleotide sequences can also be used as

CC immunogenic compositions, especially where the vector directs the

CC expression of a neutralising epitope of C. pneumoniae.

XX

SQ Sequence 20 BP; 5 A; 6 C; 4 G; 5 T; 0 other;

Query Match 1.1%; Score 20; DB 20; Length 20;

Best Local Similarity 100.0%; Pred.No. 11;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 54 ATTCTGCAAAACCGTGATGG 73

DB 20 ATTCTGCAAAACCGTGATGG 1

RESULT 8

AAK95807

ID AAK95807 standard; DNA; 20 BP.

XX

AC AAK95807;

XX

DT 13-SEP-1999 (first entry)

XX

DE PCR primer used to amplify an ORF of Chlamydia pneumoniae.

XX

KW Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;

KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;

KW vaccine; neutralising epitope; PCR primer; ss.

XX

OS Synthetic.

OS Chlamydia pneumoniae.

XX

PN WO927105-A2.

XX

PD 03-JUN-1999.

XX

PF 20-NOV-1998; 98WO-IB01890.

XX

PR 04-NOV-1998; 98US-0107078.

PR 21-NOV-1997; 97FR-0014673.

XX

PA (GEST) GENSET.

XX

PI Griffais R;

XX

DR WPI; 1999-357842/30.

XX

PT Genome sequence of Chlamydia pneumoniae

XX

PS Page 1777; Disclosure; 1912pp; English.

XX

CC AAK91991-X97517 represent PCR primers used to amplify open reading

CC frames and other nucleic acid sequences from the genome of

CC Chlamydia pneumoniae (see AAK91990). C. pneumoniae causes respiratory

CC disease such as pneumonia and bronchitis and is thought to be a

CC contributing factor in heart disease, sarcoidosis, sinusitis, purulent

CC otitis media, erythema nodosum or pharyngitis. The polypeptides encoded

CC by the open reading frames of the C. pneumoniae genome (see AAY34584-

CC AAY35879) can be used in immunogenic compositions as vaccines. Vectors

CC containing C. pneumoniae nucleotide sequences can also be used as

CC immunogenic compositions, especially where the vector directs the

CC expression of a neutralising epitope of C. pneumoniae.

ID	AA933334 standard; DNA; 20 BP.
XX	
AC	AA933334;
XX	
DT	13-SEP-1999 (first entry)
XX	
DE	PCR primer used to amplify an ORF of Chlamydia pneumoniae.
XX	
KW	Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;
KW	sinusitis; purulent otitis media; erythema nodosum; pharyngitis;
KW	vaccine; neutralising epitope; PCR primer; ss.
XX	
OS	Synthetic.
OS	Chlamydia pneumoniae.
PX	
PN	WQ9927105-A2.
XX	
PD	03-JUN-1999.
XX	
PF	20-NOV-1998; 98WO-IB01890.
XX	
PR	04-NOV-1998; 98US-0107078.
PR	21-NOV-1997; 97FR-0014673.
XX	
PA	(GEST) GENSET.
XX	
PI	Griffais R;
XX	
DR	WPI; 1999-357842/30.
XX	
PT	Genome sequence of Chlamydia pneumoniae
PS	Page 1581; Disclosure; 1912pp; English.
XX	
CC	AA931991-X97517 represent PCR primers used to amplify open reading frames and other nucleic acid sequences from the genome of Chlamydia pneumoniae (see AA931990). C. pneumoniae causes respiratory disease such as pneumonia and bronchitis and is thought to be a contributing factor in heart disease, sarcoidosis, sinusitis, purulent otitis media, erythema nodosum or pharyngitis. The polypeptides encoded by the open reading frames of the C. pneumoniae genome (see AA934584-AA935879) can be used in immunogenic compositions as vaccines. Vectors containing C. pneumoniae nucleotide sequences can also be used as immunogenic compositions, especially where the vector directs the expression of a neutralising epitope of C. pneumoniae.
SC	Sequence 20 BP; 5 A; 6 C; 4 G; 5 T; 0 other;
Query Match	1.1%; Score 20; DB 20; Length 20;
Best Local Similarity	100.0%; Pred No. 11;
Matches	20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY	953 AGCGATGGTCTCAAAATGCG 972
Db	20 AGCGATGGTCTCAAAATGCG 1
RESULT 11	
AAC05080	
ID	AAC05080 standard; cDNA; 236 BP.
XX	
AC	AAC05080;
XX	
DT	06-OCT-2000 (first entry)
XX	
DE	Human secreted protein 5' EST, SEQ ID NO: 9155.
XX	
KW	Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW	gene therapy; chromosome mapping; ss.
XX	
OS	Homo sapiens.
XX	
PN	RP1033401-A2

```

XX PD 06-SEP-2000.
XX PF 21-FEB-2000; 2000EP-0200610.
XX PR 26-FEB-1999; 99US-0122487.
XX PS (GEST ) GENSET.
XX PA Dumas Milne Edwards J, Duclert A, Giordano J;
XX PI WPI; 2000-500381/45.
XX DR MPI; 2000-500381/45.
XX PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
XX PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
XX PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX PS Claim 1; SEQ ID 9155; 71bp + CD-ROM; English.
XX CC The present sequence is one of a large number of 5' ESTs derived from
XX CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
XX CC identified within the present sequence. The 5' ESTs were prepared from
XX CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
XX CC sequences usually correspond mainly to the 3' untranslated region (UTR)
XX CC of the mRNA because they are often obtained from oligo-dT primed cDNA
XX CC libraries. Such ESTs are not well suited for isolating cDNA sequences
XX CC derived from the 5' ends of mRNAs and even in those cases where longer
XX CC cDNA sequences have been obtained, the full 5' UTR is rarely included.
XX CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
XX CC used to obtain full length cDNAs with intact 5' ends and can therefore be
XX CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.
XX CC They are used to obtain upstream regulatory sequences and to design
XX CC expression and secretion vectors.
XX SQ Sequence 236 BP; 69 A; 51 C; 43 G; 73 T; 0 other;
XX
Query Match 1.0%; Score 19; DB 21; Length 236;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 681 TTTCTCTTCATGTTTAC 699
 DB 113 TTTCTCTTCATGTTTAC 131

RESULT 12
 AAC03937
 ID AAC03937 standard; cDNA; 420 BP.

XX AC AAC03937;
 XX DT 06-OCT-2000 (first entry)
 XX DE Human secreted protein 5' EST, SEQ ID NO: 3935.
 XX KM Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
 XX KM gene therapy; chromosome mapping; ss.
 XX OS Homo sapiens.
 XX PN EP1033401-A2.
 XX PD 06-SEP-2000.
 XX PF 21-FEB-2000; 2000EP-0200610.
 XX PR 26-FEB-1999; 99US-0122487.
 XX PA (GEST) GENSET.
 XX PI Dumas Milne Edwards J, Duclert A, Giordano J;
 XX DR WPI; 2000-500381/45.

```

DR P-PSDB; AAC03931.
XX XX
XX PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
XX PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
XX PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX PS Claim 1; SEQ ID 3935; 71bp + CD-ROM; English.
XX CC The present sequence is one of a large number of 5' ESTs derived from
XX CC mRNAs encoding secreted proteins. An ORF has been identified within the
XX CC sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs
XX CC derived from 30 different tissues. EST sequences usually correspond
XX CC mainly to the 3' untranslated region (UTR) of the mRNA because they are
XX CC often obtained from oligo-dT primed cDNA libraries. Such ESTs are not
XX CC well suited for isolating cDNA sequences derived from the 5' ends of
XX CC mRNAs and even in those cases where longer cDNA sequences have been
XX CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
XX CC mRNAs with intact 5' ends and can therefore be used to obtain full length
XX CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
XX CC gene therapy and chromosome mapping procedures. They are used to obtain
XX CC upstream regulatory sequences and to design expression and secretion
XX CC vectors.
XX SQ Sequence 420 BP; 100 A; 112 C; 107 G; 101 T; 0 other;
XX
Query Match 1.0%; Score 19; DB 21; Length 420;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 370 AGCTGTTCTTCGATCTT 388
 DB 376 AGCTGTTCTTCGATCTT 394

RESULT 13
 AAC03932/c
 ID AAC03932 standard; DNA; 439 BP.

XX AC AAC03932;
 XX DT 05-NOV-2001 (first entry)
 XX DE Human brain expressed single exon probe SEQ ID NO: 3923.
 XX KM Human; brain expressed exon; gene expression analysis; probe;
 XX KM microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
 XX KM epilepsy; cancer; ss.
 XX OS Homo sapiens.
 XX PN WO200157275-A2.
 XX PD 09-AUG-2001.
 XX PF 30-JAN-2001; 2001WO-US00667.
 XX PR 04-FEB-2000; 2000US-0180312.
 XX PR 26-MAY-2000; 2000US-0207456.
 XX PR 30-JUN-2000; 2000US-0608408.
 XX PR 03-AUG-2000; 2000US-0632366.
 XX PR 21-SEP-2000; 2000US-0234687.
 XX PR 27-SEP-2000; 2000US-0236359.
 XX PR 04-OCT-2000; 2000GB-0024263.
 XX PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX DR WPI; 2001-483446/52.

Single exon nucleic acid probes for analyzing gene expression in human
 brains -

PS Example 4; SEQ ID NO: 3923; 650pp + Sequence Listing; English.

XX The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC brain. They can be used to measure gene expression in brain cell samples,
CC which may enable the diagnosis and improved treatment of nervous system
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC epilepsy and cancers. The present sequence is one of the probes of the
CC invention.

XX SQ Sequence 439 BP; 126 A; 97 C; 89 G; 127 T; 0 other;

Query Match 1.0%; Score 19; DB 22; Length 439;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 370 AGCTGTTCTTCTGAATCTT 388
Db 412 AGCTGTTCTTCTGAATCTT 394

RESULT 14
AAI13982/C
ID AAI13982 standard; DNA; 439 BP.

XX AAI13982;

XX 12-OCT-2001 (first entry)

XX Probe #3915 for gene expression analysis in human cervical cell sample.

XX Probe; human; microarray; gene expression; cervical epithelial cell;
XX cervical cancer; ss.

XX Homo sapiens.

XX WO200157278-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US00670.

XX 04-FEB-2000; 2000US-0180312.

XX 26-MAY-2000; 2000US-0207456.

XX 30-JUN-2000; 2000US-0608408.

XX 03-AUG-2000; 2000US-0632366.

XX 21-SEP-2000; 2000US-0234687.

XX 27-SEP-2000; 2000US-0236359.

XX 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488901/53.

XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human cervical epithelial cells -
XX Claim 25; SEQ ID No 3915; 487pp; English.

XX The present invention relates to human single exon nucleic acid probes
CC (SENPs). The present sequence is one such probe. The SENPs are derived
CC from human HeLa cells. The SENPs can be used to produce a single exon
CC microarray, which can be used for measuring human gene expression in a
CC sample derived from human cervical epithelial cells. By measuring gene
CC expression, the probes are therefore useful in grading and/or staging
CC of diseases of the cervix, notably cervical cancer.

XX Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 439 BP; 126 A; 97 C; 89 G; 127 T; 0 other;

Query Match 1.0%; Score 19; DB 22; Length 439;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 370 AGCTGTTCTTCTGAATCTT 388
Db 412 AGCTGTTCTTCTGAATCTT 394

RESULT 15
AAI03846/C
ID AAI03846 standard; DNA; 439 BP.

XX AAI03846;

XX 09-OCT-2001 (first entry)

XX Probe #3837 used to measure gene expression in human breast sample.

XX Probe; human; breast disease; breast cancer; development disorder; ss;
XX inflammatory disease; proliferative breast disease; non-carcinoma tumour.

XX Homo sapiens.

XX WO200157270-A2.

XX 09-AUG-2001.

XX 29-JAN-2001; 2001WO-US00661.

XX 04-FEB-2000; 2000US-0180312.

XX 26-MAY-2000; 2000US-0207456.

XX 30-JUN-2000; 2000US-0608408.

XX 03-AUG-2000; 2000US-0632366.

XX 21-SEP-2000; 2000US-0234687.

XX 27-SEP-2000; 2000US-0236359.

XX 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-476286/51.

XX Novel single exon nucleic acid probe used to measuring gene expression
XX in a human breast -
XX Claim 25; SEQ ID No 3837; 322pp; English.

XX The present invention relates to novel single exon nucleic acid probes.
CC The present sequence is one such probe. The probes are useful for
CC measuring human gene expression in a human breast sample, where the probe
CC hybridises at high stringency to a nucleic acid expressed in the human
CC breast. The probes are useful for predicting, diagnosing, grading,
CC staging, monitoring and prognosing diseases of the human breast,
CC particularly those diseases with polygenic aetiology. The diseases
CC include: breast cancer, disorders of development, inflammatory diseases
CC of the breast, fibrocystic changes, proliferative breast disease and
CC non-carcinoma tumours.

XX Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 439 BP; 126 A; 97 C; 89 G; 127 T; 0 other;

Query Match 1.0%; Score 19; DB 22; Length 439;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 370 AGCTGTTCTTCTGAATCTT 388
Db 412 AGCTGTTCTTCTGAATCTT 394

```

RESULT 16
AAK16682/c
ID AAK16682 standard; DNA; 445 BP.
XX
XX
AC AAK16682;
XX
XX
DT 05-NOV-2001 (first entry)
XX
XX
DE Human brain expressed single exon probe SEQ ID NO: 16673.
XX
XX
KW Human; brain expressed exon; gene expression analysis; probe;
KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
KW epilepsy; cancer; ss.
XX
XX
OS Homo sapiens.
XX
XX
PN MO200157275-A2.
XX
XX
PD 09-AUG-2001.
XX
XX
PF 30-JAN-2001; 2001MO-US00667.
XX
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0603408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0235359.
PR 04-OCT-2000; 2000GB-0024263.
XX
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX
PT WPI; 2001-483446/52.
XX
XX
DR WPI; 2001-483446/52.
XX
XX
PT Single exon nucleic acid probes for analyzing gene expression in human
PT brains -
XX
XX
PS Example 4; SEQ ID NO: 15673; 650bp + Sequence Listing; English.
XX
XX
CC The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC brain. They can be used to measure gene expression in brain cell samples,
CC which may enable the diagnosis and improved treatment of nervous system
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC epilepsy and cancers. The present sequence is one of the probes of the
CC invention.
XX
XX
SQ Sequence 445 BP; 119 A; 99 C; 93 G; 134 T; 0 other;
XX
XX
Query Match 1.0%; Score 19; DB 22; Length 445;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
OY 370 AGCTGTTCTTCTGATCTT 388
DB 286 AGCTGTTCTTCTGATCTT 268
XX
XX
RESULT 17
AAI23198/c
ID AAI23198 standard; DNA; 445 BP.
XX
XX
AC AAI23198;
XX
XX
DT 12-OCT-2001 (first entry)
XX
XX
DE Probe #1131 for gene expression analysis in human cervical cell sample.
XX
XX
KW Probe; human; microarray; gene expression; cervical epithelial cell;
XX

```

```

KW cervical cancer; ss.
XX
XX
OS Homo sapiens.
XX
XX
PN MO200157278-A2.
XX
XX
PD 09-AUG-2001.
XX
XX
PF 30-JAN-2001; 2001MO-US00670.
XX
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX
DR WPI; 2001-488901/53.
XX
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human cervical epithelial cells -
XX
XX
PS Claim 25; SEQ ID NO 13131; 487bp; English.
XX
XX
CC The present invention relates to human single exon nucleic acid probes
CC (SENPs). The present sequence is one such probe. The SENPs are derived
CC from human HeLa cells. The SENPs can be used to produce a single exon
CC microarray, which can be used for measuring human gene expression in a
CC sample derived from human cervical epithelial cells. By measuring gene
CC expression, the probes are therefore useful in grading and/or staging
CC of diseases of the cervix, notably cervical cancer.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at http://wipo.int/pubd/published\_pct\_sequences.
XX
XX
SQ Sequence 445 BP; 119 A; 99 C; 93 G; 134 T; 0 other;
XX
XX
Query Match 1.0%; Score 19; DB 22; Length 445;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
OY 370 AGCTGTTCTTCTGATCTT 388
DB 286 AGCTGTTCTTCTGATCTT 268
XX
XX
RESULT 18
AAI08848/c
ID AAI08848 standard; DNA; 445 BP.
XX
XX
AC AAI08848;
XX
XX
DT 09-OCT-2001 (first entry)
XX
XX
DE Probe #8839 used to measure gene expression in human breast sample.
XX
XX
KW Probe; human; breast disease; breast cancer; development disorder; ss;
KW inflammatory disease; proliferative breast disease; non-carcinoma tumour.
XX
XX
OS Homo sapiens.
XX
XX
PN MO200157270-A2.
XX
XX
PD 09-AUG-2001.
XX
XX
PF 29-JAN-2001; 2001MO-US00661.
XX
XX
PR 04-FEB-2000; 2000US-0180312.
XX

```

PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX Penn SG, Hanzel DK, Chen W, Rank DR;
 XX WPI; 2001-476286/51.
 DR Novel single exon nucleic acid probe used to measuring gene expression
 PT in a human breast -
 XX Claim 25; SEQ ID No 8839; 322pp; English.
 XX The present invention relates to novel single exon nucleic acid probes.
 CC The present sequence is one such probe. The probes are useful for
 CC measuring human gene expression in a human breast sample, where the probe
 CC hybridises at high stringency to a nucleic acid expressed in the human
 CC breast. The probes are useful for predicting, diagnosing, grading,
 CC staging, monitoring and prognosing diseases of the human breast,
 CC particularly those diseases with polygenic aetiology. The diseases
 CC include: breast cancer, disorders of development, inflammatory diseases
 CC of the breast, fibrocystic changes, proliferative breast disease and
 CC non-carcinoma tumours.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX SQ Sequence 445 BP; 119 A; 99 C; 93 G; 134 T; 0 other;
 Query Match 1.0%; Score 19; DB 22; Length 445;
 Best Local Similarity 100.0%; Pred. No. 40;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 370 AGCTGTTCTTCTGAATCTT 388
 Db 286 AGCTGTTCTTCTGAATCTT 268
 RESULT 19
 AAK50436/c
 ID AAK50436 standard; DNA; 469 BP.
 XX AAK50436;
 XX 06-NOV-2001 (first entry)
 XX Human bone marrow expressed single exon probe SEQ ID NO: 24993.
 XX Human; bone marrow expressed exon; gene expression analysis; probe;
 KW microarray; cancer; leukaemia; lymphoma; myeloma; ss.
 XX Homo sapiens.
 XX WO200157276-A2.
 XX 09-AUG-2001.
 XX 30-JAN-2001; 2001WO-US00668.
 XX 04-FEB-2000; 2000US-0180312.
 XX 26-MAY-2000; 2000US-0207456.
 XX 30-JUN-2000; 2000US-0608408.
 XX 03-AUG-2000; 2000US-0632366.
 XX 21-SEP-2000; 2000US-0234687.
 XX 27-SEP-2000; 2000US-0236359.
 XX 04-OCT-2000; 2000GB-0024263.
 XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;
 XX WPI; 2001-488900/53.
 XX Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human bone marrow -
 XX Example 4; SEQ ID NO: 24993; 658pp + Sequence Listing; English.
 XX The present invention provides a number of single exon nucleic acid
 CC probes which are derived from genomic sequences expressed in the human
 CC bone marrow. They can be used to measure gene expression in bone marrow
 CC samples, which may enable the improved diagnosis and treatment of cancers
 CC such as lymphoma, leukaemia and myeloma. The present sequence is one of
 CC the probes of the invention.
 XX SQ Sequence 469 BP; 125 A; 106 C; 98 G; 140 T; 0 other;
 Query Match 1.0%; Score 19; DB 22; Length 469;
 Best Local Similarity 100.0%; Pred. No. 40;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 370 AGCTGTTCTTCTGAATCTT 388
 Db 310 AGCTGTTCTTCTGAATCTT 292
 RESULT 20
 AAK37567/c
 ID AAK37567 standard; DNA; 561 BP.
 XX AAK37567;
 XX 06-NOV-2001 (first entry)
 XX Human bone marrow expressed single exon probe SEQ ID NO: 12124.
 XX Human; bone marrow expressed exon; gene expression analysis; probe;
 KW microarray; cancer; leukaemia; lymphoma; myeloma; ss.
 XX Homo sapiens.
 XX WO200157276-A2.
 XX 09-AUG-2001.
 XX 30-JAN-2001; 2001WO-US00668.
 XX 04-FEB-2000; 2000US-0180312.
 XX 26-MAY-2000; 2000US-0207456.
 XX 30-JUN-2000; 2000US-0608408.
 XX 03-AUG-2000; 2000US-0632366.
 XX 21-SEP-2000; 2000US-0234687.
 XX 27-SEP-2000; 2000US-0236359.
 XX 04-OCT-2000; 2000GB-0024263.
 XX (MOLE-) MOLECULAR DYNAMICS INC.
 XX Penn SG, Hanzel DK, Chen W, Rank DR;
 XX WPI; 2001-488900/53.
 XX Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human bone marrow -
 XX Example 4; SEQ ID NO: 12124; 658pp + Sequence Listing; English.
 XX The present invention provides a number of single exon nucleic acid
 CC probes which are derived from genomic sequences expressed in the human
 CC bone marrow. They can be used to measure gene expression in bone marrow
 CC samples, which may enable the improved diagnosis and treatment of cancers
 CC such as lymphoma, leukaemia and myeloma. The present sequence is one of

CC the probes of the invention.
CC
XX
SQ Sequence 561 BP; 157 A; 123 C; 114 G; 167 T; 0 other;
Query Match 1.0%; Score 19; DB 22; Length 561;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 370 AGCTGTTCTTCGAATCTT 388
|||||
Db 356 AGCTGTTCTTCGAATCTT 338
RESULT 21
AAH08684
ID AAH08684 standard; cDNA; 812 BP.
XX
AC AAH08684;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human cDNA clone (5'-primer) SEQ ID NO:5519.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
XX (HELI-) HELIX RES INST.
XX
PA Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
PS Claim 1; SEQ ID 5519; 2537bp + CD ROM; English.
XX
XX The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dr primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises at least 15 nucleotides, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632

CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
CC
XX
SQ Sequence 812 BP; 216 A; 188 C; 206 G; 198 T; 4 other;
Query Match 1.0%; Score 19; DB 22; Length 812;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 370 AGCTGTTCTTCGAATCTT 388
|||||
Db 350 AGCTGTTCTTCGAATCTT 368
RESULT 22
AAH05672
ID AAH05672 standard; cDNA; 827 BP.
XX
AC AAH05672;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human cDNA clone (5'-primer) SEQ ID NO:2507.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
XX (HELI-) HELIX RES INST.
XX
PA Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
PS Claim 1; SEQ ID 2507; 2537bp + CD ROM; English.
XX
XX The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dr primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to

CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.

XX
 SQ Sequence 827 BP; 222 A; 192 C; 192 G; 208 T; 13 other;
 Query Match 1.0%; Score 19; DB 22; Length 827;
 Best Local Similarity 100.0%; Pred. No. 41;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 370 AGCTGTTCTTCTGAATCTT 388
 |||||
 Db 358 AGCTGTTCTTCTGAATCTT 376

RESULT 23
 ABK35236
 ID ABK35236 standard; cDNA; 1150 BP.

XX
 AC ABK35236;

XX
 DT 08-MAY-2002 (first entry)

XX
 DE Human cDNA encoding secreted protein #374.

XX
 KW Human; secreted protein; gene; ss; nutritional supplement; haemophilia;
 KW viral infection; bacterial infection; fungal infection; diabetes; asthma;
 KW autoimmune disorder; rheumatoid arthritis; multiple sclerosis; tumour;
 KW autoimmune thyroiditis; allergic reaction; neurodegenerative disease;
 KW Alzheimer's disease; Parkinson's disease; liver fibrosis; cancer; ulcer;
 KW coagulation disorder; inflammatory disorder; Crohn's disease; incision;
 KW tissue regeneration; wound healing; burn; haematopoiesis;
 KW myeloid cell deficiency; lymphoid cell deficiency.

XX
 OS Homo sapiens.

XX
 PN WO200177288-A2.

XX
 PD 18-OCT-2001.

XX
 PF 29-MAR-2001; 2001WO-US10224.

XX
 PR 06-APR-2000; 2000US-195582P.

XX
 PA (GEMY) GENETICS INST INC.

XX
 PI Wong GG, Clark HF, Fachtel K, Agostino MJ, Howes SH, Resnick RJ;
 PI Gulukota K, Graham JR;

XX
 DR WPI; 2002-179321/23.

XX
 PT Five hundred and ninety two polynucleotides derived from a variety of
 PT human tissue sources which encode secreted proteins, useful for
 PT treating immune deficiencies and disorders such as autoimmune disorders

XX
 PS Claim 1; Page 268; 372pp; English.

XX
 CC The invention relates to 592 polynucleotides which have been derived from
 CC a variety of human tissue sources and which encode novel secreted
 CC proteins. The polynucleotides can be used as probes for the
 CC identification and isolation of full length cDNA and genomic DNA. The
 CC polynucleotides and proteins can also be used as nutritional supplements.
 CC The proteins are useful in the treatment of various immune deficiencies
 CC and disorders such as viral infections, bacterial infections, fungal
 CC infections, autoimmune disorders (e.g. rheumatoid arthritis, multiple
 CC sclerosis, autoimmune thyroiditis and diabetes) and allergic reactions
 CC and conditions (e.g. asthma). They are also useful for treating
 CC neurodegenerative diseases (e.g. Alzheimer's disease, Parkinson's
 CC disease), liver fibrosis, coagulation disorders (e.g. haemophilia),
 CC inflammatory disorders (e.g. Crohn's disease) and tumours. They are also
 CC useful for tissue regeneration, for wound healing and in the treatment of
 CC burns, incisions and ulcers. The proteins are also useful for regulating

CC haematopoiesis and for treating myeloid or lymphoid cell deficiencies.
 CC Sequences ABK34863-ABK35454 represent polynucleotides of the invention.

XX
 SQ Sequence 1150 BP; 326 A; 247 C; 287 G; 290 T; 0 other;

Query Match 1.0%; Score 19; DB 24; Length 1150;
 Best Local Similarity 100.0%; Pred. No. 41;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 370 AGCTGTTCTTCTGAATCTT 388
 |||||
 Db 337 AGCTGTTCTTCTGAATCTT 355

RESULT 24

AAC81029

ID AAC81029 standard; cDNA; 1187 BP.

XX
 AC AAC81029;

XX
 DT 13-FEB-2001 (first entry)

XX
 DE Human secreted protein cDNA sequence #2.

XX
 KW Cytostatic; immunosuppressive; nontropic; neuroprotective; antiviral;
 KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;
 KW vulnary; anticonvulsant; antibacterial; antifungal; antiparasitic;
 KW cardiant; Gene therapy; cancer; immune disorder; cardiovascular disorder;
 KW neurological disease; infection; human; secreted protein; ss.

XX
 OS Homo sapiens.

XX
 PN WO200063230-A2.

XX
 PD 26-OCT-2000.

XX
 PF 23-MAR-2000; 2000WO-US07677.

XX
 PR 26-MAR-1999; 99US-0126601.

XX
 PR 17-SEP-1999; 99US-0154373.

XX
 PR 14-JAN-2000; 2000US-0176064.

XX
 PA (HUMA-) HUMAN GENOME SCI INC.

XX
 PI Rosen CA, Ruben SM, Komatsoulis G;

XX
 DR WPI; 2000-647515/62.

XX
 DR P-PSDB; AAB45227.

XX
 PT New nucleic acid molecules encoding 49 human secreted proteins for
 PT diagnosing, preventing, treating or ameliorating medical conditions and
 PT used as food additives or preservatives -

XX
 PS Claim 1; Page 340; 402pp; English.

XX
 CC The present invention relates to the isolation of genes encoding

XX
 CC 49 human secreted proteins. The genes can be used to

XX
 CC generate fusion proteins by linking to the gene for the human

XX
 CC immunoglobulin G Fc portion for increasing the stability of

XX
 CC the fusion protein as compared to the human protein only. The genes and

XX
 CC proteins are useful for preventing, ameliorating or treating medical

XX
 CC conditions, e.g. by protein or gene therapy. The genes are isolated

XX
 CC from a range of human tissues disclosed in the specification. The

XX
 CC nucleic acids, proteins, antibodies and (ant)agonists are useful in

XX
 CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast

XX
 CC and ovarian cancer, and other cancers of the adrenal gland, bone, bone

XX
 CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;

XX
 CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune

XX
 CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's

XX
 CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative colitis;

XX
 CC (c) cardiovascular disorders such as myocardial ischaemias; (d) wound

XX
 CC healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy;

XX
 CC and (f) infectious diseases such as viral, bacterial, fungal and

CC parasitic infections. The present sequence is a cDNA encoding a protein
CC of the invention.
XX
SQ Sequence 1187 BP, 374 A, 252 C, 271 G, 288 T, 2 other;
Query Match 1.0%; Score 19; DB 21; Length 1187;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 370 AGCTGTTCTTCTGAATCTT 388
Db 354 AGCTGTTCTTCTGAATCTT 372
RESULT 25
AA158527
ID AA158527 standard; cDNA; 1240 BP.
XX
AC AA158527;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 730.
XX
XX Human; neurotrophic; immunosuppressant; cytostatic; gene therapy; cancer;
XX peripheral nervous system; neuropathy; central nervous system; CNS;
XX Alzheimer's; Parkinson's disease; Huntington's disease; hemostatic;
XX amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemoketic;
XX chemoketic; thrombolytic; drug screening; arthritis; inflammation;
XX leukemia; ss.
XX
OS Homo sapiens.
XX
PN WO200153312-A1.
XX
PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000WO-US31263.
XX
PR 21-JAN-2000; 2000US-0483725.
XX
PR 25-APR-2000; 2000US-0552317.
XX
PR 09-JUL-2000; 2000US-0593042.
XX
PR 19-JUL-2000; 2000US-0620312.
XX
PR 03-AUG-2000; 2000US-0653450.
XX
PR 14-SEP-2000; 2000US-0662191.
XX
PR 19-OCT-2000; 2000US-0693036.
XX
PR 29-NOV-2000; 2000US-0727344.
XX
XX
PA (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
XX Wang J, Wang Z, Wehman T, Xu C, Xue AJ, Yang Y, Zhang J;
XX Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
DR WPI: 2001-442253/47.
XX
DR P-PSDB; AAM39371.
XX
PT Novel nucleic acids and polypeptides, useful for treating disorders
XX such as central nervous system injuries -
XX
PS Claim 1; SEQ ID NO 730; 10078bp; English.
XX
XX The invention relates to human nucleic acids (AA157798-AA161369) and
XX the encoded polypeptides (AAM38642-AA42213) with neurotrophic,
XX immunosuppressant and cytostatic activity. The polynucleotides are useful
XX in gene therapy. A composition containing a polypeptide or polynucleotide
XX of the invention may be used to treat diseases of the peripheral nervous
XX system, such as peripheral nervous injuries, peripheral neuropathy and
XX localised neuropathies and central nervous system diseases, such as
XX Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
XX lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
XX utilisation of the activities such as: Immune system suppression,
XX Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic

CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
SQ Sequence 1240 BP, 351 A, 272 C, 306 G, 311 T, 0 other;
Query Match 1.0%; Score 19; DB 22; Length 1240;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 370 AGCTGTTCTTCTGAATCTT 388
Db 420 AGCTGTTCTTCTGAATCTT 438
RESULT 26
AAH33729
ID AAH33729 standard; cDNA; 1311 BP.
XX
XX AAH33729;
XX
DT 03-SEP-2001 (first entry)
XX
DE Human colon cancer antigen encoding cDNA SEQ ID NO:785.
XX
XX Human; colon cancer; colon cancer antigen; diagnosis; detection;
XX colorectal carcinoma; ss.
XX
XX Homo sapiens.
XX
OS Homo sapiens.
XX
PN WO200122920-A2.
XX
PD 05-APR-2001.
XX
PF 28-SEP-2000; 2000WO-US26524.
XX
PR 29-SEP-1999; 99US-0157137.
XX
PR 03-NOV-1999; 99US-0163280.
XX
XX (HDMA-) HDMA GENOME SCT INC.
XX
XX Ruben SM, Barash SC, Birse CE, Rosen CA;
XX
XX WPI: 2001-235357/24.
XX
XX P-PSDB; AAG74298.
XX
PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
XX useful for preventing, diagnosing and/or treating colorectal cancers -
XX
XX Claim 1; Page 2755; 9803bp; English.
XX
XX AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
XX cancer-associated nucleic acid molecules (N) and proteins (P), where
XX the proteins are collectively known as colon cancer antigens. The colon
XX cancer antigens have cytostatic activity and can be used in gene
XX therapy and vaccine production. N and P may be used in the prevention,
XX diagnosis and treatment of diseases associated with inappropriate P
XX expression. For example, N and P may be used to treat disorders
XX associated with decreased expression by rectifying mutations or deletions
XX in a patient's genome that affect the activity of P by expressing
XX inactive proteins or to supplement the patient's own production of P.
XX Additionally, N may be used to produce the colon cancer-associated Ps,
XX by inserting the nucleic acids into a host cell and culturing the cell
XX to express the proteins. N and P can be used in the prevention, diagnosis
XX and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
XX and AAG77789 represent sequences used in the exemplification of the
XX present invention.
XX
XX N.B. Pages 666 to 682 and page 7053 of the sequence listing were
XX missing at time of publication, meaning no sequences are present for
XX SEQ ID NO:1027 to 1052, 7921 and 7922.

SQ Sequence 1311 BP; 349 A; 282 C; 316 G; 361 T; 3 other;
Query Match 1.0%; Score 19; DB 22; Length 1311;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 370 AGCTGTTCTTCTGAATCTT 388
Db 387 AGCTGTTCTTCTGAATCTT 405

RESULT 27
AAI60313/C
ID AAI60313 standard; cDNA; 1684 BP.
XX AC AAI60313;
XX DT 22-OCT-2001 (first entry).
XX DE Human polynucleotide SEQ ID NO 4302.
XX KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia; ss.
XX OS Homo sapiens.
XX PN WO200153312-A1.
XX PD 26-JUL-2001.
XX PF 26-DEC-2000; 2000WO-US34263.
XX PR 21-JAN-2000; 2000US-0488725.
XX PR 25-APR-2000; 2000US-0552317.
XX PR 09-JUL-2000; 2000US-0598042.
XX PR 19-JUL-2000; 2000US-0620312.
XX PR 03-AUG-2000; 2000US-0653450.
XX PR 14-SEP-2000; 2000US-0662191.
XX PR 19-OCT-2000; 2000US-0693036.
XX PR 29-NOV-2000; 2000US-0727344.
XX PA (HYSE-) HYSEQ INC.
XX PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX WPI; 2001-442253/47.
XX P-PSDB; AAM41157.
XX PT Novel nucleic acids and polypeptides, useful for treating disorders
XX PT such as central nervous system injuries -
XX PS Claim 1; SEQ ID NO 4302; 10078pp; English.
XX CC The invention relates to human nucleic acids (AAI57798-AAI61369) and
XX CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
XX CC immunosuppressant and cytostatic activity. The polynucleotides are useful
XX CC in gene therapy. A composition containing a polypeptide or polynucleotide
XX CC of the invention may be used to treat diseases of the peripheral nervous
XX CC system, such as peripheral nervous injuries, peripheral neuropathy and
XX CC localised neuropathies and central nervous system diseases, such as
XX CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
XX CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
XX CC utilisation of the activities such as: Immune system suppression,
XX CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
XX CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
XX CC assays for receptor activity, arthritis and inflammation, leukaemias and
XX CC C.N.S disorders.

CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX SQ Sequence 1684 BP; 485 A; 388 C; 358 G; 453 T; 0 other;
Query Match 1.0%; Score 19; DB 22; Length 1684;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 370 AGCTGTTCTTCTGAATCTT 388
Db 1346 AGCTGTTCTTCTGAATCTT 1328

RESULT 28
AAH68588
ID AAH68588 standard; cDNA; 1684 BP.
XX AC AAH68588;
XX DT 13-SEP-2001 (first entry)
XX DE Human protein HP10522 coding sequence.
XX KW Human; gene therapy; tumour; ss.
XX OS Homo sapiens.
XX PN WO200142302-A1.
XX PD 14-JUN-2001.
XX PF 06-DEC-2000; 2000WO-JP08631.
XX PR 06-DEC-1999; 99JP-0346863.
XX PR 08-DEC-1999; 99JP-0346864.
XX PR 08-FEB-2000; 2000JP-0031062.
XX PR 10-FEB-2000; 2000JP-0034090.
XX PR 14-FEB-2000; 2000JP-0034091.
XX PR 14-FEB-2000; 2000JP-0035829.
XX PR 14-FEB-2000; 2000JP-0035899.
XX PR 14-MAR-2000; 2000JP-0071161.
XX PR 30-MAY-2000; 2000JP-0160851.
XX PA (NISC-) JAPAN SCI & TECHNOLOGY CORP.
XX PI Kato S, Eguchi C, Saeki M;
XX WPI; 2001-381646/40.
XX P-PSDB; AAG93303.
XX PT Human protein originated from tumor cell line, applicable as drug,
XX PT reagent for studying intracellular protein networks and protein source
XX PT for drug screening, also encoded cDNA for gene diagnosis and gene
XX PT therapy -
XX PS Claim 3; Pages 349-353; 471pp; Japanese.
XX CC The present sequence is a human protein coding sequence. The human
XX CC protein, preferably originated from tumor cell line, is applicable as a
XX CC drug, a reagent for studying intracellular protein networks and a protein
XX CC source for screening proteins for binding low molecular weight drugs. The
XX CC human protein coding sequence is useful for gene diagnosis and gene
XX CC therapy, expression vectors and transformant cells for detection of
XX CC ligands and receptors.
XX SQ Sequence 1684 BP; 444 A; 360 C; 390 G; 490 T; 0 other;
Query Match 1.0%; Score 19; DB 22; Length 1684;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 370 AGCTGTTCTTCTGAATCTT 388

DB 351 AGCTGTTCTTCGAACTT 369

RESULT 29

AAH16819
ID AAH16819 standard; cDNA; 1685 BP.

AC AAH16819;

DT 26-JUN-2001 (first entry)

DE Human cDNA sequence SEQ ID NO:16081.

DE Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

OS Homo sapiens.

PN EP1074617-A2.

PD 07-FEB-2001.

PF 28-JUL-2000; 2000EP-0115126.

PR 29-JUL-1999; 99JP-0243036.

PR 27-AUG-1999; 99JP-0300253.

PR 11-JAN-2000; 2000JP-0113776.

PR 02-MAY-2000; 2000JP-0183767.

PR 09-JUN-2000; 2000JP-0241899.

XX (HELI-) HELIX RES INST.

XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX WPI; 2001-318749/34.

PS Claim 8; SEQ ID 16081; 2537pp + CD ROM; English.

XX The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.

XX Sequence 1685 BP; 442 A; 362 C; 392 G; 489 T; 0 other;

Query Match 1.0%; Score 19; DB 22; Length 1685;

Best Local Similarity 100.0%; Pred. No. 42;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 370 AGCTGTTCTTCGAACTT 388
DB 350 AGCTGTTCTTCGAACTT 368

RESULT 30

AAH16007
ID AAH16007 standard; cDNA; 1690 BP.

AC AAH16007;

DT 26-JUN-2001 (first entry)

DE Human cDNA sequence SEQ ID NO:14646.

DE Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

OS Homo sapiens.

PN EP1074617-A2.

PD 07-FEB-2001.

PF 28-JUL-2000; 2000EP-0116126.

PR 29-JUL-1999; 99JP-0248036.

PR 27-AUG-1999; 99JP-0300253.

PR 11-JAN-2000; 2000JP-0118776.

PR 02-MAY-2000; 2000JP-0183767.

PR 09-JUN-2000; 2000JP-0241899.

XX (HELI-) HELIX RES INST.

XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX WPI; 2001-318749/34.

PS Claim 8; SEQ ID 14646; 2537pp + CD ROM; English.

XX The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.

XX Sequence 1690 BP; 444 A; 361 C; 390 G; 495 T; 0 other;

Query Match 1.0%; Score 19; DB 22; Length 1690;

Best Local Similarity 100.0%; Pred. No. 42;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 370 AGCTGTTCTTCTGAATCTT 388
 Db 357 AGCTGTTCTTCTGAATCTT 375

RESULT 31
 AAD24382
 ID AAD24382 standard; cDNA; 1749 BP.
 XX
 AC AAD24382;
 XX
 DT 07-MAR-2002 (first entry)
 XX
 DE Human RNA metabolism protein-9 (RMEP-9) cDNA.
 XX
 KW Human; RNA metabolism protein-9; RMEP-9; gout; nervous system disorder;
 KW auto-immune; inflammatory; cell proliferative; developmental; thyroiditis;
 KW gene therapy; epilepsy; dementia; stroke; Alzheimer's disease; amnesia;
 KW Parkinson's disease; prion disease; insomnia; endocrine disorder; AIDS;
 KW Acquired Immune Deficiency Syndrome; mental disorder; allergy; anaemia;
 KW asthma; atherosclerosis; Crohn's disease; rheumatoid arthritis; vaccine;
 KW glomerulonephritis; multiple sclerosis; diabetes mellitus; osteoporosis;
 KW cancer; cirrhosis; hepatitis; psoriasis; transgenic animal; anti-ulcer;
 KW tranquiliser; drug screening; pancreatitis; renal tubular acidosis;
 KW systemic lupus erythematosus; colitis; ss.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 48..1046
 FT /*tag= a
 FT /product= "Human RMEP-9 protein"
 FT sig_peptide 48..137
 FT /*tag= b
 FT mat_peptide 138..1043
 FT /*tag= c
 FT /product= "Mature human RMEP-9 protein"
 FT
 XX WO200183524-A2.
 XX
 XX 08-NOV-2001.
 XX
 XX 27-APR-2001; 2001WO-US13862.
 XX
 XX 28-APR-2000; 2000US-200184P.
 XX 04-MAY-2000; 2000US-201875P.
 XX 04-MAY-2000; 2000US-202090P.
 XX 06-JUN-2000; 2000US-210232P.
 XX 25-JUL-2000; 2000US-220553P.
 XX (INCY-) INCYTE GENOMICS INC.
 XX
 XX Lal P, Yue H, Tang YT, Lu DAM, Azimzai Y, Au-Young J, Hillman JL;
 XX Baughn MR, Yao MG, Burford N, Batra S, Policky JU;
 XX
 XX WPI; 2002-034502/04.
 XX P-PSDB; AAE15246.
 XX
 XX New human RNA metabolism protein for diagnosing or treating nervous
 XX system disorders, autoimmune/inflammatory disorders, cell proliferative
 XX disorders and developmental disorders -
 XX
 XX Claim 11; Page 175; 196pp; English.
 XX
 XX The invention relates to human RNA metabolism proteins (RMEP) and their
 XX corresponding cDNA molecules. RMEP and its DNA are used for diagnosing,
 XX treating and preventing nervous system disorders (epilepsy, dementia,
 XX stroke, Alzheimer's disease, Huntington's disease, Parkinson's disease);
 XX prion diseases; fatal familial insomnia, nutritional and metabolic
 XX diseases of the nervous system; inherited, metabolic, endocrine and
 XX toxic myopathy; mental disorders (mood, anxiety, schizophrenic disorders)
 XX amnesia and Tourette's disorder; autoimmune/inflammatory disorders (AIDS-

CC acquired immune deficiency syndrome, allergies, anaemia, asthma, gout,
 CC atherosclerosis, Crohn's disease, diabetes mellitus, glomerulonephritis,
 CC Hashimoto's thyroiditis, multiple sclerosis, osteoarthritis, rheumatoid
 CC arthritis, osteoporosis, pancreatitis, systemic lupus erythematosus,
 CC ulcerative colitis, and infections); cell proliferative disorders (cancer
 CC arteriosclerosis, cirrhosis, hepatitis, psoriasis); and developmental
 CC disorders (renal tubular acidosis). RMEP DNA is useful in drug screening.
 CC techniques, gene therapy and for creating transgenic animals. The present
 CC sequence is human RMEP-9 cDNA.
 XX

SQ Sequence 1749 BP; 474 A; 373 C; 396 G; 506 T; 0 other;

Query Match 1.0%; Score 19; DB 24; Length 1749;
 Best Local Similarity 100.0%; Pred. No. 42;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 370 AGCTGTTCTTCTGAATCTT 388
 Db 386 AGCTGTTCTTCTGAATCTT 404

RESULT 32
 AAV99969/c
 ID AAV99969 standard; cDNA; 2364 BP.
 XX
 AC AAV99969;
 XX

DT 07-JUN-1999 (first entry)
 XX
 DE Nucleotide sequence of the human smoothened gene.
 XX

KW Human; smoothened gene; smoothened peptides; SPS; liver repair;
 KW cell proliferation; cell differentiation; spermatogenesis;
 KW osteogenesis; chondrogenesis; apoptosis; degenerative disorder;
 KW neoplastic disorder; hyperplastic disorder; varicella-zoster;
 KW tumour; ds.
 XX

OS Homo sapiens.

FH Key Location/Qualifiers
 FT CDS 1..2364
 FT /*tag= a
 FT /product= "smoothened polypeptide"
 FT

XX WO9901468-A2.

XX 14-JAN-1999.

XX 01-JUL-1998; 98WO-US13793.

XX 21-JUL-1997; 97US-0897798.

XX 01-JUL-1997; 97US-0051407.

XX (INGH/) INGHAM P W.
 XX (ONTO-) ONTOGENY INC.
 XX (VHEU/) VAN DEN HEUVEL M.

XX WPI; 1999-105996/09.

XX P-PSDB; AAW81064.

XX New isolated smoothened genes - used to develop products for
 XX treating e.g. cancers, neurodegenerative disorders, nervous system
 XX injury, immunological diseases or infections

XX Claim 12; Pages 90-91; 107pp; English.

XX This is the nucleotide sequence of the novel human smoothened
 CC gene used in the method of the invention to produce smoothened
 CC polypeptides (SPs). The SPs can modulate at least one of
 CC proliferation, differentiation or survival of a cell which
 CC expresses the SP. The products can be used to modulate
 CC spermatogenesis, osteogenesis, chondrogenesis or neuronal cell
 CC differentiation or to enhance survival of neuronal cells, e.g. to

CC prevent apoptosis. They can be used for treating e.g.
 CC degenerative disorders marked by loss of particular cell-types,
 CC apoptosis, neoplastic and/or hyperplastic disorders. In particular
 CC they can be used to treat neurological conditions derived from
 CC acute, subacute, or chronic injury to the nervous system, aging of
 CC the nervous system, chronic neurodegenerative diseases of the
 CC nervous system, chronic immunological diseases of the nervous
 CC system or affecting the nervous system. They can also treat be used
 CC to CNS trauma infection, infection (such as viral infection with
 CC varicella-zoster), metabolic disease, nutritional deficiency,
 CC toxic agents (such as cisplatin treatment), chronic pain syndromes
 CC or in nerve prostheses for the repair of central and peripheral
 CC nerve damage, malignant gliomas, medulloblastomas, neuroectodermal
 CC tumours, ependymomas and for liver repair and regeneration of lung
 CC tissue in the treatment of emphysema, or for enhancing bond
 CC formation, tissue formation or fertility. The products can also be
 CC used for detection, diagnosis and drug screening.

XX Sequence 2364 BP; 412 A; 764 C; 721 G; 467 T; 0 other;

Query Match 1.0%; Score 19; DB 20; Length 2364;

Best Local Similarity 100.0%; Pred. No. 43;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1280 TTGGCAATCATCTTGCTCT 1298

DB 1712 TTGGCAATCATCTTGCTCT 1694

RESULT 33
 AAL39688/c
 ID AAL39688 standard; cDNA; 2364 BP.

XX AAL39688;

XX 05-SEP-2002 (first entry)

XX cDNA encoding the human smoothened Drosophila homologue (SMOH).

XX Cytostatic; polymorphic variant; single nucleotide polymorphism; SMOH;

KM human smoothened Drosophila homologue; basal cell carcinoma; BCC;

XX gene therapy; antisense gene therapy; gene; ss.

XX Homo sapiens.

XX Key

XX CDS

Location/Qualifiers
 1..2364
 /*tag= a
 /product= "Protein of human smoothened Drosophila
 homologue gene"
 replace(74,G)
 /*tag= b
 replace(384,T)
 /*tag= c
 replace(528,G)
 /*tag= d
 replace(582,G)
 /*tag= e
 replace(621,T)
 /*tag= f
 replace(672,T)
 /*tag= g
 replace(852,A)
 /*tag= h
 replace(1939,T)
 /*tag= i
 replace(2052,A)
 /*tag= j
 replace(2259,A)
 /*tag= k

PD 11-APR-2002.
 XX
 XX 04-OCT-2001; 2001WO-US31304.
 XX
 XX 04-OCT-2000; 2000US-237871P.
 PR
 XX (GENA-) GENA15SANCE PHARM INC.
 PA
 XX Bentivegna SC, Choi JY, Koshiy B, Lee HH, Sausker EA;
 PI
 XX WPI; 2002-519113/55.
 DR
 XX P-PSDB; AAO21722.
 XX
 PT New genetic variants of smoothened Drosophila homolg (SMOH) gene
 PT useful for therapeutic purposes and for expressing SMOH protein useful
 PT in identifying drugs to treat basal cell carcinomas
 PS
 XX Claim 25; Fig 2; 179pp; English.

CC The invention relates to an isolated polynucleotide comprising a sequence
 CC which is a polymorphic variant of a reference sequence for the human
 CC smoothened Drosophila homologue (SMOH) gene or its fragment, or a
 CC polymorphic variant of a reference sequence for a SMOH cDNA or its
 CC fragment. A new isolated polypeptide is useful for screening for drugs
 CC targeting the polypeptide. A new method is useful for identifying an
 CC association between a trait such as a clinical response to a drug
 CC targeting SMOH and a haplotype or haplotype pair of SMOH gene. The
 CC methods have applicability in developing diagnostic tests and therapeutic
 CC treatments for basal cell carcinomas (BCCs). The isolated polynucleotide
 CC is useful for studying the expression and function of SMOH and expressing
 CC SMOH protein for use in screening for candidate drugs to treat diseases
 CC related to SMOH activity. The polymorphism and haplotype data are useful
 CC for validating whether SMOH is a suitable target for drugs to treat BCCs,
 CC screening for the drugs and reducing bias in clinical trials of the
 CC drugs. The isolated polynucleotide is useful for therapeutic purposes.
 CC The new method, an oligonucleotide and kit of the invention are useful
 CC for determining whether an individual has one of the haplotypes or the
 CC haplotype pairs. The polynucleotides of the invention can be used to
 CC treat disorders by gene therapy and antisense gene therapy. This
 CC polynucleotide sequence represents the cDNA encoding the human smoothened
 CC Drosophila homologue (SMOH) of the invention.

XX Sequence 2364 BP; 412 A; 764 C; 721 G; 467 T; 0 other;

Query Match 1.0%; Score 19; DB 24; Length 2364;

Best Local Similarity 100.0%; Pred. No. 43;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1280 TTGGCAATCATCTTGCTCT 1298

DB 1712 TTGGCAATCATCTTGCTCT 1694

RESULT 34
 AAV29125/c
 ID AAV29125 standard; cDNA; 2972 BP.

XX AAV29125;

XX 11-SEP-1998 (first entry)

XX Homo sapiens Smoothened protein gene.

XX Smoothened; treatment; cancer; vertebrate; vSmo; tissue;

KM regeneration; skin; lung; muscle; neural; bone; gut; disease;

XX ageing; trauma; detection; diagnosis; drug screening; ss.

XX Homo sapiens.

XX Key

XX CDS

Location/Qualifiers
 13..2376
 /*tag= a
 /product= Smoothened protein

XX WO200229004-A2.

```

XX PN WO9814475-A1.
XX PD 09-APR-1998.
XX PF 29-SEP-1997; 97WO-US17433.
XX PR 30-SEP-1996; 96US-0720484.
XX PA (GETH ) GENENTECH INC.
XX PI De SAUVAGE FJ, Rosenthal A, Stone DM;
XX DR WPI; 1998-240021/21.
XX DR P-PSDB; AAW37912.
XX PT New isolated vertebrate Smoothed proteins - is used to develop
XX PT products for treatment of cancers or for use in inducing
XX PT regeneration of tissue damaged by disease, ageing or trauma
XX PS Claim 14; Fig 4; 58pp; English.
XX CC The sequence is that coding for human Smoothed protein (Smo).
XX CC Smo interacts with Hedgehog and Patched signalling molecules
XX CC involved in cell proliferation and differentiation. The products
XX CC can be used to develop agents to provide protection from conditions
XX CC associated with constitutive activity of Smo or Hedgehog, including
XX CC some forms of cancer that may result from e.g. basal cell carcinoma,
XX CC basal cell nevus syndrome and pancreatic carcinoma. Smo agonists
XX CC can be used to induce the formation of, or enhance or stimulate
XX CC tissue regeneration, such as regeneration of skin tissue, lung
XX CC tissue, muscle (such as heart or skeletal muscle), neural tissue
XX CC (such as serotonergic neurons, motoneurons or straital neurons),
XX CC bone tissue or gut tissue, which may be damaged by disease,
XX CC ageing or trauma. The products can also be used for detection,
XX CC diagnosis and drug screening.
XX SQ Sequence 2972 BP; 525 A; 929 C; 918 G; 600 T; 0 other;

Query Match 1.0%; Score 19; DB 19; Length 2972;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1280 TTGGCAATCATCTTGCTCT 1298
Db 1724 TTGGCAATCATCTTGCTCT 1706

RESULT 35
AAC65391/c
AC AAC65391 standard; cDNA; 2972 BP.
AC AAC65391;
XX 13-FEB-2001 (first entry)
DE Human Smoothed cDNA.
KW Human; Smoothed; hSmo; cell proliferation; cell differentiation;
KW cell signalling; ss.
XX Homo sapiens.
XX OS US6136958-A.
XX PN 24-OCT-2000.
XX PD 30-SEP-1997; 97US-0953823.
XX PF 30-SEP-1996; 96US-0027070.
XX PR (GETH ) GENENTECH INC.
XX PA

Stone DM, De Sauvage FJ, Rosenthal A;
WPI; 2001-006229/01.
P-PSDB; AAB28646.
Antibodies to vertebrate smoothed proteins which relate to Hedgehog
and patched signaling molecules involved in cell proliferation and
differentiation -
Example 3; Fig 4; 38pp; English.
The present sequence encodes human Smoothed. Smoothed interacts with
Hedgehog and Patched signalling molecules involved in cell proliferation
and differentiation. The rat Smoothed cDNA was isolated from an
embryonic day 9-10 rat cDNA library. The rat Smoothed cDNA was then
used as a probe to isolate the human homologue from a human embryonic
lung cDNA library. Antibodies which specifically bind to the vertebrate
smoothed polypeptides were identified.
Sequence 2972 BP; 525 A; 929 C; 918 G; 600 T; 0 other;

Query Match 1.0%; Score 19; DB 22; Length 2972;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1280 TTGGCAATCATCTTGCTCT 1298
Db 1724 TTGGCAATCATCTTGCTCT 1706

RESULT 36
ABL32627
ID ABL32627 standard; DNA; 19082 BP.
XX ABL32627;
AC ABL32627;
XX 26-MAR-2002 (first entry)
DT Human immune system associated gene SEQ ID NO: 600.
DE Human; immune system disease; cytosine methylation; antiaesthetic;
KW antiarteriosclerotic; antianaemic; cytosine; neurotropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
KW gene; ds.
XX Homo sapiens.
XX OS WO200200928-A2.
XX PN 03-JAN-2002.
XX PD 02-JUL-2001; 2001WO-EP07537.
XX PF 30-JUN-2000; 2000DE-1032529.
XX PR 01-SEP-2000; 2000DE-1043826.
XX PA (EPIG-) EPIGENOMICS AG.
XX PI Olek A, Piepenbrock C, Berlin K;
XX DR WPI; 2002-130909/17.
XX PT Nucleic acid comprising fragment of chemically modified gene, useful
XX PT for diagnosis and treatment of diseases associated with abnormal
XX PT cytosine methylation -
XX PS Claim 1; SEQ ID NO 600; 32pp + Sequence Listing; German.
XX CC The present invention provides a number of human immune system associated

```

CC genes which are modified by the methylation of cytosines. The sequences
 CC can be used in the diagnosis and treatment of immune system disorders,
 CC including eye diseases such as retinopathy, neovascular glaucoma and
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
 CC diseases. The present sequence is a gene of the invention.

SQ Sequence 19082 BP; 5624 A; 429 C; 4025 G; 9004 T; 0 other;

Query Match 1.0%; Score 19; DB 24; Length 19082;
 Best Local Similarity 100.0%; Pred. No. 47;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 509 CGAGAGCAATTTTATTG 527
 DB 11533 CGAGAGCAATTTTATTG 11551

RESULT 37

ABL09700 ABL09700 standard; cDNA; 22660 BP.

AC ABL09700;

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 23582.

KW Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical; gene; ss.

OS Drosophila melanogaster.

PN WO200171042-A2.

PD 27-SEP-2001.

PF 23-MAR-2001; 2001WO-US09231.

PR 23-MAR-2000; 2000US-191537P.

PR 11-JUL-2000; 2000US-0611150.

PA (PEKE) PE CORP NY.

PI Venter JC, Adams M, Li PWD, Myers EW;

DR WPI; 2001-656860/75.

DR P-PSDB; ABB55597.

PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -

PS Claim 1; SEQ ID NO 23582; 21bp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1030 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
 CC sequences (AB101840-AB16175) and the encoded proteins
 CC (AB857737-AB872072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 22660 BP; 6703 A; 4609 C; 4758 G; 6590 T; 0 other;

Query Match 1.0%; Score 19; DB 23; Length 22660;
 Best Local Similarity 100.0%; Pred. No. 47;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 266 AAAAACAATGTACACAG 284
 DB 19690 AAAAACAATGTACACAG 19708

RESULT 38
 AAL39687/c
 ID AAL39687 standard; DNA; 31352 BP.

AC AAL39687;

DT 05-SEP-2002 (first entry)

DE Genomic DNA encoding the human smoothened Drosophila homologue (SMO).

KW Cytostatic; polymorphic variant; single nucleotide polymorphism; SMOH;

KW human smoothened Drosophila homologue; basal cell carcinoma; BCC;

KW gene therapy; antisense gene therapy; gene; ds.

OS Homo sapiens.

FX Location/Qualifiers

FT 4044..27353

FT /tag= a

FT /product= "Protein of human smoothened Drosophila

FT homologue gene"

FT /note= "This coding region contains 11 introns"

FT 4044..4374

FT /tag= b

FT /number= 1

FT 4375..18279

FT /tag= c

FT /number= 1

FT 18280..18485

FT /tag= d

FT /number= 2

FT 18486..20104

FT /tag= e

FT /number= 2

FT 20105..20314

FT /tag= f

FT /number= 3

FT 20315..20511

FT /tag= g

FT /number= 3

FT 20512..20684

FT /tag= h

FT /number= 4

FT 20685..21051

FT /tag= i

FT /number= 4

FT 21052..21271

FT /tag= j

FT /number= 5

FT 21272..21365

FT /tag= k

FT /number= 5

FT 21366..21489

FT /tag= l

FT /number= 6

FT 21490..23660

FT /tag= m

FT /number= 6

FT 23661..23753

FT /tag= n

FT /number= 7

FT 23754..24190

FT /tag= o

FT /number= 7

FT 24191..24299

FT /tag= p

FT /number= 8

FT exon

FT intron

FT intron 24300..25264
 FT /tag= q
 FT /number= 8
 FT 25265..25450
 FT /tag= r
 FT /number= 9
 FT 25451..25866
 FT /tag= s
 FT /number= 9
 FT 25867..26015
 FT /tag= t
 FT /number= 10
 FT 26016..26537
 FT /tag= u
 FT /number= 10
 FT 26538..26672
 FT /tag= v
 FT /number= 11
 FT 26673..26925
 FT /tag= w
 FT /number= 11
 FT 26926..27353
 FT /tag= x
 FT /number= 12
 FT replace(4117,G)
 FT /tag= y
 FT replace(4440,C)
 FT /tag= z
 FT replace(18224,C)
 FT /tag= aa
 FT replace(18332,T)
 FT /tag= ab
 FT replace(18476,G)
 FT /tag= ac
 FT replace(20014,A)
 FT /tag= ad
 FT replace(20079,C)
 FT /tag= ae
 FT replace(20149,G)
 FT /tag= af
 FT replace(20188,T)
 FT /tag= ag
 FT replace(20239,T)
 FT /tag= ah
 FT replace(20338,G)
 FT /tag= ai
 FT replace(20358,T)
 FT /tag= aj
 FT replace(20360,C)
 FT /tag= ak
 FT replace(20386,G)
 FT /tag= al
 FT replace(20616,A)
 FT /tag= am
 FT replace(20752,A)
 FT /tag= an
 FT replace(20763,C)
 FT /tag= ao
 FT replace(21284,A)
 FT /tag= ap
 FT replace(24368,G)
 FT /tag= aq
 FT replace(26068,T)
 FT /tag= ar
 FT replace(26133,A)
 FT /tag= as
 FT replace(26146,C)
 FT /tag= at
 FT replace(26492,T)
 FT /tag= au
 FT replace(26856,A)
 FT /tag= av
 FT replace(26867,C)

FT variation /tag= aw
 FT replace(26928,T)
 FT /tag= ax
 FT replace(27041,A)
 FT /tag= ay
 FT replace(27248,A)
 FT /tag= az
 FT replace(27442,T)
 FT /tag= ba
 FT replace(27572,A)
 FT /tag= bb
 XX WO200229004-A2.
 XX 11-APR-2002.
 XX PD
 XX 04-OCT-2001; 2001WO-US31304.
 XX PF
 XX 04-OCT-2000; 2000US-237871P.
 XX PR
 XX (GENA-) GENAISSANCE PHARM INC.
 XX PA
 XX Bentivegna SC, Choi JY, Koshy B, Lee HH, Sausker EA;
 XX PI
 XX WPI; 2002-519113/55.
 XX DR
 XX P-PSDB; AAO21722.
 XX
 FT New genetic variants of smoothened Drosophila homolog (SMOH) gene
 FT useful for therapeutic purposes and for expressing SMOH protein useful
 FT in identifying drugs to treat basal cell carcinomas -
 XX
 PS Claim 1; Fig 1; 179pp; English.
 XX
 CC The invention relates to an isolated polynucleotide comprising a sequence
 CC which is a polymorphic variant of a reference sequence for the human
 CC smoothened Drosophila homologue (SMOH) gene or its fragment, or a
 CC polymorphic variant of a reference sequence for a SMOH cDNA or its
 CC fragment. A new isolated polypeptide is useful for screening for drugs
 CC targeting the polypeptide. A new method is useful for identifying an
 CC association between a trait such as a clinical response to a drug
 CC targeting SMOH and a haplotype or haplotype pair of SMOH gene. The
 CC methods have applicability in developing diagnostic tests and therapeutic
 CC treatments for basal cell carcinomas (BCCs). The isolated polynucleotide
 CC is useful for studying the expression and function of SMOH and expressing
 CC SMOH protein for use in screening for candidate drugs to treat diseases
 CC related to SMOH activity. The polymorphism and haplotype data are useful
 CC for validating whether SMOH is a suitable target for drugs to treat BCCs,
 CC screening for the drugs and reducing bias in clinical trials of the
 CC drugs. The isolated polynucleotide is useful for therapeutic purposes.
 CC The new method, an oligonucleotide and kit of the invention are useful
 CC for determining whether an individual has one of the haplotypes or the
 CC haplotype pairs. The polynucleotides of the invention can be used to
 CC treat disorders by gene therapy and antisense gene therapy. This
 CC polynucleotide sequence represents the genomic DNA encoding the human
 CC smoothened Drosophila homologue (SMOH) of the invention.
 XX
 SQ Sequence 31352 BP; 7528 A; 7628 C; 7963 G; 8233 T; 0 other;
 Query Match 1.0%; Score 19; DB 24; Length 31352;
 Best Local Similarity 100.0%; Pred. No. 48;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1280 TTGGCAATCATCTTGCTCT 1298
 |||||
 Db 25926 TTGGCAATCATCTTGCTCT 25908
 RESULT 39
 ABN26694
 ID ABN26694 standard; cDNA; 235 BP.
 XX
 AC ABN26694;
 XX

DT 24-JUN-2002 (first entry)

XX Human ORFX polynucleotide sequence SEQ ID NO:21865.

DE Human: open reading frame; ORFX; gene therapy; cancer; cirrhosis;

XX hyperproliferative disorder; psoriasis; benign tumor; haemorrhage;

KW degenerative disorder; osteoarthritis; neurodegenerative disorder;

KW cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;

KW hypertension; hypothyroidism; cholesterol ester storage disease;

KW immune deficiency; immune disorder; infectious disease;

KW autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;

KW myasthenia gravis; gene; ss.

XX Homo sapiens.

OS WO200192523-A2.

PN 06-DEC-2001.

XX 29-MAY-2001; 2001WO-US10836.

XX 30-MAY-2000; 2000US-206132P.

PR 29-AUG-2000; 2000US-228716P.

XX (CURA-) CURAGEN CORP.

XX Shimketa RA, Leach MD;

PI WPI; 2002-106308/14.

DR P-PSDB; ABP10942.

XX Novel human polypeptides and polynucleotides useful for diagnosing,

PT preventing and treating cardiovascular disease, neurodegenerative,

PT hyperproliferative disorders and autoimmune disorders -

XX Disclosure; SEQ ID 21865; 1037pp; English.

XX The present invention describes substantially purified human proteins

CC (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1

CC in the specification)). ABN15762 to ABN27252 encode the human ORFX

CC proteins given in ABP00010 to ABP1500. ORFX proteins are useful for

CC treating or preventing a pathology associated with an ORFX-associated

CC disorder in humans, and in the manufacture of a medicament for treating a

CC syndrome associated with ORFX-associated disorder. ORFX polynucleotide

CC sequences can be used in gene therapy. ORFX sequences can be used in the

CC treatment of cancer, hyperproliferative disorders, cirrhosis of liver,

CC psoriasis, benign tumors, keloid, degenerative disorders, haemorrhage,

CC osteoarthritis, neurodegenerative disorders, disorders related to organ

CC transplantation, cardiovascular diseases, diabetes mellitus, systemic

CC lupus erythematosus, hypertension, hypothyroidism, cholesterol ester

CC storage disease, various immune deficiencies and disorders, infectious

CC diseases, autoimmune disorders such as multiple sclerosis, rheumatoid

CC arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host

CC disease and autoimmune inflammatory eye disease. ORFX proteins are also

CC useful for treating burns, incisions, ulcers, for treating osteoporosis,

CC bone degenerative disorders, or periodontal disease, and for gut

CC protection or regeneration and treatment of lung or liver fibrosis,

CC reperfusion injury in various tissues and conditions resulting from

CC systemic cytokine damage.

CC N.B. The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 235 BP; 64 A; 42 C; 62 G; 67 T; 0 other;

SQ

Query Match 1.0%; Score 18; DB 24; Length 235;

Best Local Similarity 100.0%; Pred. No. 1.3e+02;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 709 CATGACTAAATTTTCTT 726

DB 120 CATGACTAAATTTTCTT 137

RESULT 40

AA191598/C

ID AA191598 standard; CDNA; 357 BP.

XX AA191598;

AC 06-NOV-2001 (first entry)

XX 06-NOV-2001 (first entry)

DT Human polynucleotide SEQ ID NO 11658.

XX Human: cytokine; cell proliferation; cell differentiation; gene therapy;

KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;

KW tissue growth factor; immunomodulatory; cancer; leukaemia;

KW nervous system disorders; arthritis; inflammation; ss.

XX Homo sapiens.

OS WO200164835-A2.

PN 07-SEP-2001.

XX 26-FEB-2001; 2001WO-US04927.

XX 28-FEB-2000; 2000US-0515126.

PR 18-MAY-2000; 2000US-0577409.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Drmanac RT;

PI WPI; 2001-514838/56.

DR P-PSDB; AA011667.

XX Isolated nucleic acids and polypeptides, useful for preventing

PT diagnosing and treating e.g. leukaemia, inflammation and immune

PT disorders -

XX Claim 1; SEQ ID NO 11658; 1399pp + Sequence Listing; English.

PS The invention relates to human polynucleotides (AA179941-AA193841) and

XX the encoded proteins (AA000010-AA013910) that exhibit activity elating to

CC cytokine, cell proliferation or cell differentiation or which may induce

CC production of other cytokines in other cell populations. The vaccines or

CC polynucleotides and polypeptides are useful in gene therapy, vaccines or

CC peptide therapy. The polypeptides have various cytokine-like activities,

CC e.g. stem cell growth factor activity, haematopoiesis regulating

CC activity, tissue growth factor activity, immunomodulatory activity and

CC activity/inhibit activity and may be useful in the diagnosis and/or

CC treatment of cancer, leukaemia, nervous system disorders, arthritis and

CC inflammation.

CC Note: The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 357 BP; 115 A; 61 C; 88 G; 92 T; 1 other;

SQ

Query Match 1.0%; Score 18; DB 22; Length 357;

Best Local Similarity 100.0%; Pred. No. 1.3e+02;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 110 AGCCCTAATTTTAAACGTC 127

DB 257 AGCCCTAATTTTAAACGTC 240

RESULT 41

ABN90719

ID ABN90719 standard; DNA; 378 BP.

XX ABN90719;

AC 24-JUL-2002 (first entry)

XX Staphylococcus epidermidis ORF nucleic acid sequence SEQ ID NO:182.
DE Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;
XX antibacterial; gene therapy; gene; ds.
KW Staphylococcus epidermidis.
XX US6380370-B1.
PN 30-APR-2002.
XX 13-AUG-1998; 98US-0134001.
XX 14-AUG-1997; 97US-055779P.
PR 08-NOV-1997; 97US-064964P.
XX (GENO-) GENOME THERAPEUTICS CORP.
XX Doucette-Stamm LA, Bush D;
PI WPI; 2002-381255/41.
XX P-PSDB; ABP38174.
XX Novel isolated nucleic acid encoding a Staphylococcus epidermidis
PT polypeptide, useful for diagnosing and treating bacterial infections -
FT Disclosure; SEQ ID 182; 267pp; English.
XX ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading
CC frame (ORF) nucleic acid sequences which encode the amino acid sequences
CC given in ABP5124 to ABP37960. The S. epidermidis sequences have
CC antibacterial activity and can be used in gene therapy. The sequences
CC can also be used in the diagnosis and treatment of bacterial infections,
CC particularly S. epidermidis infections. The sequences can be used to
CC screen for compounds able to interfere with the S. epidermidis life
CC cycle or inhibit S. epidermidis infection.
CC N.B. The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC USPTO web site.
XX SQ Sequence 378 BP; 150 A; 57 C; 65 G; 106 T; 0 other;

Query Match 1.0%; Score 18; DB 24; Length 378;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 266 AAAAAACAATGTACACAA 283
Db ||||||||||||||||
211 AAAAAACAATGTACACAA 228

RESULT 42
ABK44466/C
ID ABR44466 standard; cDNA; 459 BP.
XX AC ABK44466;
XX 05-JUN-2002 (first entry)
XX cDNA encoding colon tumour protein, SEQ ID No 17.
XX Human; colon tumour; vaccine; colon cancer; immunogenic;
KW immunotherapy; gene; ss.
XX Homo sapiens.
XX WO200212328-A2.
PN 14-FEB-2002.
XX 31-JUL-2001; 2001WO-US24218.

XX Staphylococcus epidermidis ORF nucleic acid sequence SEQ ID NO:182.
DE Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;
XX antibacterial; gene therapy; gene; ds.
KW Staphylococcus epidermidis.
XX US6380370-B1.
PN 30-APR-2002.
XX 13-AUG-1998; 98US-0134001.
XX 14-AUG-1997; 97US-055779P.
PR 08-NOV-1997; 97US-064964P.
XX (GENO-) GENOME THERAPEUTICS CORP.
XX Doucette-Stamm LA, Bush D;
PI WPI; 2002-381255/41.
XX P-PSDB; ABP38174.
XX Novel isolated nucleic acid encoding a Staphylococcus epidermidis
PT polypeptide, useful for diagnosing and treating bacterial infections -
FT Disclosure; SEQ ID 182; 267pp; English.
XX ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading
CC frame (ORF) nucleic acid sequences which encode the amino acid sequences
CC given in ABP5124 to ABP37960. The S. epidermidis sequences have
CC antibacterial activity and can be used in gene therapy. The sequences
CC can also be used in the diagnosis and treatment of bacterial infections,
CC particularly S. epidermidis infections. The sequences can be used to
CC screen for compounds able to interfere with the S. epidermidis life
CC cycle or inhibit S. epidermidis infection.
CC N.B. The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC USPTO web site.
XX SQ Sequence 378 BP; 150 A; 57 C; 65 G; 106 T; 0 other;

Query Match 1.0%; Score 18; DB 24; Length 378;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 266 AAAAAACAATGTACACAA 283
Db ||||||||||||||||
211 AAAAAACAATGTACACAA 228

RESULT 42
ABK44466/C
ID ABR44466 standard; cDNA; 459 BP.
XX AC ABK44466;
XX 05-JUN-2002 (first entry)
XX cDNA encoding colon tumour protein, SEQ ID No 17.
XX Human; colon tumour; vaccine; colon cancer; immunogenic;
KW immunotherapy; gene; ss.
XX Homo sapiens.
XX WO200212328-A2.
PN 14-FEB-2002.
XX 31-JUL-2001; 2001WO-US24218.

PR 03-AUG-2000; 2000US-223283P.
PR 28-MAR-2001; 2001US-279763P.
PR 29-JUN-2001; 2001US-302051P.
XX (CORI-) CORIXA CORP.
XX King GE, Meagher MJ, Xu J, Secrist H;
XX WPI; 2002-241739/29.
XX New colon cancer polypeptides and polynucleotides, useful as vaccines,
PT for diagnosing, preventing, and treating colon cancer, and as markers
PT for the progression of cancer -
XX Claim 1; SEQ ID No 17; 147pp; English.
XX The invention relates to polynucleotides encoding colon tumour proteins.
CC The polynucleotides and encoded polypeptides are useful in pharmaceutical
CC compositions, such as vaccines, for the diagnosis, prevention, and
CC treatment of colon cancer. Polynucleotide sequences may be used as
CC hybridisation probes or primers, and in the design and preparation of
CC ribozyme molecules for inhibiting expression of tumour polypeptides and
CC proteins in tumour cells. The compositions are useful for stimulating an
CC immune response against cancer, particularly for the immunotherapy of
CC colon cancer, and as markers for the progression of cancer.
CC ABK44450-ABK46237 represent coding sequences of human colon tumour
CC proteins of the invention.
CC Note: With the exception of SEQ ID No 1 and 2, the sequence data
CC for this patent did not form part of the printed specification but was
CC supplied by the European Patent Office.
XX SQ Sequence 459 BP; 112 A; 93 C; 107 G; 147 T; 0 other;

Query Match 1.0%; Score 18; DB 24; Length 459;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 887 TCCGTTTGGGAATGGAT 904
Db ||||||||||||||||
36 TCCGTTTGGGAATGGAT 19

RESULT 43
ABL93238/C
ID ABL93238 standard; cDNA; 478 BP.
XX AC ABL93238;
XX 10-JUN-2002 (first entry)
XX Arabidopsis thaliana nucleic acid sequence Ref:2027003 SEQ ID NO:3.
DE Arabidopsis thaliana; insecticide; fungicide; plant; mapping; diagnosis;
KW genetic modification; gene; ss.
XX Arabidopsis thaliana.
XX US2002023280-A1.
XX 21-FEB-2002.
XX 26-JAN-2001; 2001US-0770444.
XX 27-JAN-2000; 2000US-178502P.
XX (GORL/) GORLACH J.
XX (ANY/) AN Y.
XX (HAMI/) HAMILTON C M.
XX (PRIC/) PRICE J L.
XX (RAIN/) RAINES T M.
XX (YUY/) YU Y.
XX (RAME/) RAMEAKA J G.
XX (PAGE/) PAGE A.

PA (MATH/) MATHEW A. V.
PA (LEDE/) LEDFORD B. L.
PA (WOES/) WOESSNER J. P.
PA (HAAS/) HAAS W. D.
PA (GARC/) GARCIA C. A.
PA (KRICK/) KRICKER M.
PA (SLAT/) SLATER T.
PA (DAVI/) DAVIS K. R.
PA (ALLE/) ALLEN K.
PA (HOFF/) HOFFMAN N.
PA (HURB/) HURBAN P.
XX
XX Gorlach J., An Y., Hamilton CM, Price JL, Raines TM, Yu Y,
PI Ramewak JG, Page A, Mathew AV, Ledford BL, Woessner JP, Haas WD,
PI Garcia CA, Kricker M, Slater T, Davis KR, Allen K, Hoffman N,
PI Hurban P;
XX
XX WPI; 2002-267486/31.
XX
XX New Arabidopsis thaliana nucleic acid, for identifying homologous
PT gene, producing compositions that modulate the expression or function
PT of its encoded protein, and mapping functional regions of a protein -
XX
XX Claim 1; SEQ ID 3; 44pp; English.
XX
XX The present invention describes an Arabidopsis thaliana nucleic acid (I)
CC comprising a sequence capable of hybridising under stringent conditions
CC to a sequence (S1) selected from any one of the 999 sequences given in
CC AB93236 to AB94234. (I) have insecticide and fungicide activities, and
CC they can be used as protein expression modulators. (I) can be used in
CC identifying homologous or related genes, in producing compositions that
CC modulate the expression or function of their encoded proteins, mapping
CC functional regions of the proteins, and in studying associated
CC physiological pathways. (I) can also be used: (1) for the genetic
CC manipulation of cells, particularly plant cells; (2) in screening assays
CC of various plant strains to determine the strains that are best capable
CC of withstanding a particular disease or environmental stress; (3) for
CC enhancing or inhibiting production of a biosynthetic product in a plant;
CC (4) as probes in mapping and in diagnosis, in genetic modification and
CC for screening purposes, to generate additional copies of the nucleic
CC acids, to generate ribozymes or antisense oligonucleotides, and as
CC single-stranded DNA probes or as triple-strand forming oligonucleotides;
CC and (5) for generating genetically modified transgenic organisms.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC USPTO web site.
XX
XX Sequence 478 BP; 131 A; 93 C; 116 G; 138 T; 0 other;
SQ
XX
XX Query Match 1.0%; Score 18; DB 24; Length 478;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 835 TGGAGTGTCTTTCTTTT 852
DB 97 TGGAGTGTCTTTCTTTT 80
XX
XX RESULT 44
ABV19019
ID ABV19019 standard; cDNA; 576 BP.
XX
XX AC ABV19019;
XX
XX DT 13-SEP-2002 (first entry)
XX
XX DE Human prostate expression marker cDNA 19010.
XX
XX KM Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
XX pharmacogenomic marker; gene; ss.
XX
XX OS Homo sapiens.
XX

PN W0200160860-A2.
XX
XX 23-AUG-2001.
XX
XX PD
XX
XX PF 20-FEB-2001; 2001WO-US05171.
XX
XX PR 17-FEB-2000; 2000US-183119P.
XX
XX PR 16-MAR-2000; 2000US-189862P.
XX
XX PR 25-MAY-2000; 2000US-207454P.
XX
XX PR 09-JUN-2000; 2000US-211314P.
XX
XX PR 18-JUL-2000; 2000US-219007P.
XX
XX PR 13-DEC-2000; 2000US-255281P.
XX
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
PA
XX
XX Schlegel R, Endege WO, Monahan JE;
PI
XX
XX WPI; 2001-662795/76.
XX
XX DR
XX
XX PT Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer, stage of prostate cancer -
XX
XX PS Claim 1; Page 3121; 11750pp; English.
XX
XX The invention relates to an isolated nucleic acid molecule (I) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (I) is useful for:
CC (a) assessing whether a patient is afflicted with prostate cancer;
CC (b) monitoring the progression of prostate cancer in a patient;
CC (c) assessing the efficacy of a test compound to inhibit prostate
CC cancer in a patient;
CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer
CC in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound;
CC (g) determining whether prostate cancer has metastasised in a patient;
CC (h) assessing the aggressiveness or indolence of prostate cancer in a
CC patient;
CC (I) is also useful as a pharmacodynamic or pharmacogenomic marker.
XX
XX Sequence 576 BP; 189 A; 89 C; 104 G; 174 T; 20 other;
SQ
XX
XX Query Match 1.0%; Score 18; DB 23; Length 576;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 895 GGGATGGATTAAAGG 912
DB 418 GGGATGGATTAAAGG 435
XX
XX RESULT 45
ABQ66275
ID ABQ66275 standard; DNA; 613 BP.
XX
XX AC ABQ66275;
XX
XX DT 21-AUG-2002 (first entry)
XX
XX DE Arabidopsis thaliana polynucleotide SEQ ID NO 852.
XX
XX KM Arabidopsis thaliana, thale cress; plant; transgenic; GMO; disease;
XX stress; metabolic pathway; biosynthetic pathway; nutrition; fungicide;
XX insecticide; antibiotic; ds.
XX
XX OS Arabidopsis thaliana.
XX
XX PN US2002059663-A1.
XX
XX PD 16-MAY-2002.
XX
XX PF 26-JUN-2001; 2001US-0770149.

```

XX 27-JAN-2000; 2000US-178506P.
XX
XX (GORL/) GORLACH J.
XX (ANY/) AN Y.
XX (HAMI/) HAMILTON C M.
XX (PRIC/) PRICE J L.
XX (RAIN/) RAINES T M.
XX (YUY/) YU Y.
XX (RAME/) RAMEAKA J G.
XX (PAGE/) PAGE A.
XX (MATH/) MATHAW A V.
XX (LEDF/) LEDFORD B L.
XX (WOES/) WOESSNER J P.
XX (HAAS/) HAAS W D.
XX (GARC/) GARCIA C A.
XX (KRIC/) KRICKER M.
XX (SLAT/) SLATER T.
XX (DAVI/) DAVIS K R.
XX (ALLE/) ALLEN K.
XX (HOFF/) HOFFMAN N.
XX (HURB/) HURBAN P.
XX
XX Gorlach J, An Y, Hamilton CM, Price JL, Raines TM, Yu Y;
PI Rameaka JG, Page A, Mathew AV, Ledford BL, Woessner JP, Haas WD;
PI Garcia CA, Kricker M, Slater T, Davis KR, Allen K, Hoffman N;
PI Hurban P;
XX
XX WPI; 2002-479224/51.
XX
XX New nucleic acid that hybridizes to Arabidopsis thaliana sequences,
PT useful e.g. for preparing transgenic plants with increased resistance
PT or altered metabolism
XX
XX Claim 1; SEQ ID NO 852; 40pp + Sequence Listing; English.
XX
XX The invention relates to nucleic acids (I) that hybridise under stringent
CC conditions to any of 999 sequences (ABQ65424-ABQ66422) or their
CC fragments. (I) are used to express the corresponding polypeptides (II) or
CC to produce genetically modified plant cells or transgenic plants, which
CC may have improved resistance to disease or stress, or altered
CC metabolic/biosynthetic pathways (for production of commercial,
CC nutritional or medicinal products), or generally any trait of interest,
CC or can be used to screen for biologically active agents (e.g. fungicides,
CC insecticides and antibiotics).
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC USPTO at seqdata.uspto.gov/sequence.html?DocID=999909770149.
XX
XX SQ Sequence 613 BP; 177 A; 146 C; 115 G; 175 T; 0 other;
XX
XX Query Match 1.0%; Score 18; DB 24; Length 613;
XX Best Local Similarity 100.0%; Pred. No. 1.3e+02;
XX Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX Qy 517 ATTTTATTGTCCTCAAG 534
XX |||||||
XX Db 73 ATTTTATTGTCCTCAAG 90

```

Search completed: December 17, 2002, 09:05:10
Job time : 405 secs

THIS PAGE BLANK (USPTO)

